Signal processing on social media: Theory and evidence from financial markets

Peiran Jiao André Veiga Ansgar Walther*

This version: November 21, 2016

Abstract

We analyze the processing of information from social media and news media, using a unique dataset on financial markets. We find patterns consistent with a theory of social media as an "echo chamber": Social networks repeat information, but boundedly rational investors interpret repeated signals as new information. This is based on the empirical finding that stocks with high social media coverage experience *high* subsequent volatility and trading activity, while high news media coverage predicts *low* volatility and trading activity. Alternative mechanisms based on private information, investor disagreement, uncertainty shocks, and other behavioral biases are not consistent with the data.

JEL: D83, G02.

Keywords: Social media, news media, behavioral economics, asset prices.

*Jiao: Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK (e-mail: peiran.jiao@economics.ox.ac.uk); Veiga: Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK (e-mail: andre.veiga@economics.ox.ac.uk); Walther: Warwick Business School, University of Warwick, Scarman Road, Coventry CV4 7AL, UK (e-mail: ansgar.walther@wbs.ac.uk). We thank MarketPsych for providing access to the dataset. We are grateful to Daniel Andrei, Thierry Foucault, Basile Grassi, Dan Greenwald, Robin Greenwood, Gur Huberman, Ryoko Ito, Marcin Kacperczyk, Joel Peress, Dan Quigley, Tarun Ramadorai, Selma Walther, Mungo Wilson, Peyton Young and seminar participants at the CEPR European Summer Symposium in Financial Markets for their comments.

1 Introduction

In a recent survey by the Reuters Institute (2016), 51% of respondents use social media to access news every week, and 12% cite it as their main source of news.[1] A growing economic literature compares online and offline news (e.g. Gentzkow and Shapiro, 2011), but open questions remain regarding social media: How is social media content processed? Is it processed differently from traditional online and offline news? Which models best describe the role of information from different sources? In this paper, we address these questions by analyzing how financial markets respond to social and news media, using a unique dataset of stock-level news coverage from a wide range of sources.

Financial markets offer a natural testing ground for competing theories of signal processing. Our strategy is to extend standard models of optimal investment to obtain empirical predictions about the relationship between social and news media, market prices and trading activity. As a benchmark, consider a Bayesian model where social and news media deliver (similarly) informative signals and agents correctly update their beliefs. This model predicts that social media and news media coverage, as measured by the number of signals from each source, have qualitatively similar effects on subsequent prices and trading because both convey genuine information.

As an alternative, we develop a model which captures the fact that a large proportion of social media content is repeated via "sharing" or "re-tweeting". In this model, social media is an "echo chamber" which replicates existing signals. Rational agents ignore repetitions, but some boundedly rational agents treat repeated signals as if they were new information. This form of bounded rationality is motivated by theories of social networks based on persuasion bias (DeMarzo et al., 2003),[2] empirical evidence showing that financial markets react to repeated signals (Huberman and Regev, 2001; Tetlock, 2011), and by psychological evidence on repetition-induced learning (Hawkins and Hoch, 1992). We show that social media and news media coverage have *opposite* effects on market activity in this model. News media, containing genuine information, dampens disagreement between rational and boundedly rational agents and therefore decreases subsequent return volatility and trading activity. Social media, by contrast, increases disagreement and boosts the confidence of boundedly rational agents, thus increasing subsequent return volatility and trading volumes.

In the data, we find striking differences between how social and news media relate to market activity. We quantify coverage on social and news media using measures of "buzz" for approximately 3000 stocks from the Thomson Reuters MarketPsych Indices (TRMI) database. Buzz is defined as the number of stock-specific phrases or words in news articles

[1] A significant share of social media content relates to economics and finance, as witnessed by an emerging industry that extracts and sells market-relevant indicators from social media. See, for example: http://www.wsj.com/articles/tweets-give-birds-eye-view-of-stocks-1436128047

[2] In DeMarzo et al. (2003), persuasion bias denotes a similar deviation from rationality to our echo chamber: Agents communicate with each other and fail to correctly account for repetitions of information.

or social media posts, relative to aggregate coverage of all stocks. These measures are constructed from a broad spectrum of news media sources, as well as the most popular social media content. We merge these data with stock prices, trading activity as measured by turnover, and stock characteristics at the monthly frequency. Our key empirical finding is that high social media buzz around a given stock predicts statistically significant *increases* in idiosyncratic return volatility and trading activity over the following month, while high news media buzz predicts the opposite, i.e. significant *decreases* in volatility and trading activity. Neither result is driven by unobserved stock characteristics or time trends. These patterns are inconsistent with the Bayesian model, but consistent with the echo chamber hypothesis.

We then estimate a vector autoregression model and find that an increase in news media buzz leads to an increase in subsequent social media buzz in the sense of Granger causality. The converse is not true; increases in social media have no significant impact on subsequent news. These findings lend further support to the echo chamber hypothesis, where social media primarily responds to and repeats existing news.

We also consider several alternative mechanisms, none of which can match our finding that social media buzz predicts increased volatility of returns. First, we analyze a model where news media produces public signals, but social media produces private signals which are only observed by a subset of "well-connected" investors. Second, we consider models where social media is informative, but where its interpretation deviates from Bayesian updating. A key advantage of our approach is that we can derive the testable predictions of many biases which are recognized by behavioral economics: Namely, we consider *overconfident* investors who overestimate the precision of their information as in Scheinkman and Xiong (2003); *conservative* investors who are wedded to their prior beliefs as in Barberis et al. (1998); investors who cannot process all available information due to *rational inattention* as in Sims (2003); and investors whose perception of information is distorted by *confirmation bias* as in Rabin and Schrag (1999). A common feature in these models, and in the model with private information, is that social media conveys genuine information in principle. As a result, an increase in the number of social media signals tends to reduce subsequent volatility, which is at odds with our empirical findings.[3]

Finally, we consider whether our findings can be explained by the endogeneity of media coverage. For instance, an exogenous increase in underlying volatility might increase the demand for information, thus making investors more willing to participate in social media.

[3] We emphasize that this argument is not at odds with the common intuition that information increases volatility because prices are sensitive to information, especially in models with higher order uncertainty Kondor (2012). This intuition describes the volatility of current prices, while we focus on the volatility of *subsequent* returns. While prices become more more variable when information arrives, subsequent returns tend to become *less* variable because prices become more closely aligned with future fundamentals. We discuss these effects in detail in Section 3.2.

Alternatively, investors might post their opinions on social media if they disagree with one another, or after placing orders so as to "talk their own book". These are important caveats, but we present evidence to the contrary, based on the predictions of formal models with endogenous buzz: First, if buzz were driven by shocks to disagreement, then one would expect the predictive power of social media buzz for volatility to diminish once we control for measures of disagreement such as the dispersion of analyst opinions. We do not find empirical support for this prediction. Second, if buzz were driven by underlying uncertainty and the demand for information, then one would expect heightened activity in both social media and news media to reflect this demand. This is inconsistent with our finding that news and social media predict opposite movements in volatility and trading activity.

In summary, our results strongly suggest that social media content is processed in a fundamentally different way from news media, and that the data are consistent with a theory of social media as an "echo chamber". The crucial fact in our model is that the repetition of information on social media leads to more volatile beliefs about the state of the world, and to larger deviations of these beliefs from fundamental values. This mechanism has implications beyond financial markets: Echo chambers can potentially create more volatile political cycles, if electoral contests are covered extensively on social media, or larger business cycle fluctuations, if households and businesses use social media to inform consumption and investment decisions. Thus, our results suggest scope for future research into the volatility of agents' beliefs and decisions as a function of social media coverage.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3 contains our theoretical framework. In Section 4, we describe our data on news media, social media and financial variables, and present summary statistics. In Section 5, we show our main empirical results and check their robustness. In Section 6, we interpret these results in the context of our theoretical framework. Section 7 concludes.

2 Literature

Our findings relate to an existing literature on news, information processing, and financial markets. Many papers suggest that market reactions to news can deviate from Bayesian prescriptions. For instance, Cutler et al. (1989) show that macroeconomic and political news do not explain a large portion of stock market movements. Moreover, investors are prone to overreact to new information (De Bondt and Thaler, 1985), to react to "stale news" which merely repeat previous revelations (Tetlock, 2011), and to focus on "attention-grabbing" stocks rather than considering all available information (Barber and Odean, 2008). More generally, sentiments in news and online searches predict stock returns and trading volumes (Tetlock, 2007), stocks with low coverage have higher returns (Fang and Peress, 2009), and press coverage reduces information asymmetries (Bushee et al., 2010). Beyond traditional news media, noise levels in trading pits predict high volatility (Coval and Shumway, 2001), activity in specialist chat rooms (e.g. *RagingBull*) predicts high volatility and trading volume (Antweiler and Frank, 2004), and sentiment indicators extracted from online forums and searches can predict returns (Chen et al., 2014; Da et al., 2015). Our contribution in this context is twofold: First, we *directly* contrast the effects of social and news media and point out the stark differences between the two. Second, we relate our empirical results to theoretical predictions, and demonstrate which models of information processing are consistent with the evidence.

Our findings further relate to the literature on media information processing in social networks. DeMarzo et al. (2003) show that "persuasion bias", i.e. the failure to account for repetitions in available information, provides disproportionate social influence for well-connected players in social networks. Bayesian theories of social networks, on the other hand, predict convergence of beliefs to the truth under relatively mild conditions (Acemoglu et al., 2011). These papers focus on the diffusion of information on networks. In this paper, we abstract from network graphs and focus on testing the implications of Bayesian and non-Bayesian information processing for aggregate asset prices. Our contribution is to show that a non-Bayesian model, particularly one in which social media content is processed subject to persuasion bias, matches the data more easily than the Bayesian alternative.

Recent research on the economics of the media has addressed the relationship between online and offline news media content. For example, online and offline news tend to be substitutes (Gentzkow, 2007) and there is evidence of moderate segregation of online news consumption along ideological lines (Gentzkow and Shapiro, 2011; Falck et al., 2014). We contribute to this literature by offering a direct comparison between news and social media. Our results suggest that the substitutability of social media and news media is limited, because they appear to be processed in fundamentally different ways.

Finally, we contribute to the behavioral asset pricing literature, which Hirshleifer (2001) and Barberis and Thaler (2003) review in detail. Existing theoretical papers analyze the behavior of markets where investors are overconfident (Kyle and Wang, 1997; Odean, 1998; Daniel et al., 1998; Gervais and Odean, 2001; Scheinkman and Xiong, 2003), conservative (Barberis et al., 1998) and subject to confirmation bias (Rabin and Schrag, 1999; Pouget et al., 2014). Our model of information processing offers a parsimonious way of determining the empirical predictions of several behavioral biases. Moreover, we place more emphasis on investor heterogeneity than the existing literature, which allows us to derive new predictions about trading activity and return volatility.

3 Theoretical framework

We consider an asset market at two dates, $t \in \{1, 2\}$. At date 1, a unit mass of investors trade a risk-free asset with return r and a risky asset (say, a stock) with market price p. The stock is in zero net supply and yields a random payoff $\theta \sim \mathcal{N}\left(\theta_0, \rho_0^{-1}\right)$ per unit at date 2. All investors have constant absolute risk aversion (CARA) utility with parameter γ.

Our goal is to develop empirical predictions about the relationship between news media coverage, social media coverage, and financial market activity. We begin with a theory of social media as an "echo chamber" which repeats existing information. Below, we consider a range of alternative mechanisms, including the Bayesian benchmark which we discussed in the introduction.

3.1 Social media as an echo chamber

News media conveys information, summarized by N informative signals $(s_1, ..., s_i, ..., s_N)$. Each of these signals is of the form $s_i = \theta + \epsilon_i$ and the errors are normally distributed $\epsilon_i \sim \mathcal{N}(0, \rho_\epsilon^{-1})$, independently of each other and of θ. The parameter ρ_ϵ measures the precision of these signals. We think of these informative signals as being broadcast by traditional news media, so N will capture "news media buzz".

Social media repeats news, for example through re-tweets or through the sharing of news articles and videos. We assume that each informative signal is repeated $K-1$ times within the fold of social media, for a total of NK signals. The repetitions create a vector of $N(K-1)$ signals $(s_{N+1}, s_{N+2}, ..., s_{NK})$, where $s_{N+ik} \equiv s_i$ for all $i = 1, ..., N$ and all $k = 1, ..., K-1$. The effective precision of these repeated signals is zero. Thus, "social media buzz" is captured by the number of repetitions K.[4]

Investors have a common prior belief that the stock's payoff θ is normally distributed according to the true distribution $\mathcal{N}(\theta_0, \rho_0^{-1})$. At date 1, all investors observe the vector of NK public signals $\mathbf{s} = (s_1, ..., s_i, ..., s_{NK})$.

A mass $1 - \lambda$ of investors are *rational*. After observing the public signals $\mathbf{s}$, rational investors believe that $\theta|\mathbf{s} \sim \mathcal{N}(\theta_R, \rho_R^{-1})$, where

$$\theta_R = \sum_{i=1}^{NK} w_i s_i + \left(1 - \sum_{i=1}^{NK} w_i\right)\theta_0, \tag{1}$$

$$\rho_R = \rho_0 + N\rho_\epsilon. \tag{2}$$

The rational weights for updating in equation 1 are $w_i = \rho_\epsilon/(\rho_0 + N\rho_\epsilon)$ when $i \leq N$, and $w_i = 0$ when $i > N$. Rational investors recognize that social media signals are repetitions

[4]In this model, the total number of social media signals is $(K-1)\,N$, so an increase in news buzz N also increases activity on social media. Our predictions are unchanged in a model where each informative signal is observed K/N times, so that news and social media buzz move independently.

and ignore them.

A mass λ of investor are *behavioral*. Behavioral investors treat social media signals as if they were new information, and thus believe that $\theta|\mathbf{s} \sim \mathcal{N}(\theta_B, \rho_R^{-1})$, where

$$\theta_B = \sum_{i=1}^{NK} \hat{w}_i s_i + \left(1 - \sum_{i=1}^{NK} \hat{w}_i\right) \theta_0, \tag{3}$$

$$\rho_B = \rho_0 + NK\rho_\epsilon. \tag{4}$$

Behavioral investors interpret all signals, including their repetitions on social media, as informative signals with precision ρ_ϵ.[5] Consequently, they attach weight $\hat{w}_i = \rho_\epsilon/(\rho_0 + NK\rho_\epsilon)$ to all signals.

If stocks trade at price p at date 0, rational investors will demand $x_R(p) = \rho_R\gamma^{-1}(\theta_R - rp)$ units, and behavioral investors will demand $x_B(p) = \rho_B\gamma^{-1}(\theta_B - rp)$ units. The equilibrium price $p^\star$ solves $x_B + x_R = 0$, or equivalently,

$$rp^\star = q\theta_B + (1 - q)\theta_R, \tag{5}$$

where the parameter $q = \lambda\rho_B/(\lambda\rho_B + (1 - \lambda)\rho_R)$ is a *confidence-weighted* measure of behavioral investors.

3.2 Empirical predictions

We now derive predictions about the impact of news and social media "buzz" on the subsequent volatility of risky asset returns and trading activity. The volatility of returns between dates 1 and 2 is the variance $\mathbb{V}[R] = \mathbb{V}[\theta/p^\star]$. The return R is the ratio of two Gaussian random variables θ and $p^\star$, and there is no closed-form solution for its variance (Hinkley, 1969). In this section, we use an analytical approximation for return volatility to derive testable predictions. Moreover, we confirm that our predictions match the outcomes of numerically simulated markets where $\mathbb{V}[R]$ is computed exactly. To approximate $\mathbb{V}[R]$, note that the beliefs of behavioral investors satisfy the law of iterated expectations: $\mathbb{E}[\theta_B] = \theta_0$: Behavioral investors over-react to social media, but this reaction is equally strong for positive and negative signals, so that their average mistake is zero. Then, the equilibrium price matches the prior on average: $\mathbb{E}[rp^\star] = \theta_0$. Using a first-order approximation around the unconditional mean, we obtain

$$\frac{\theta_0^2}{r^2}\mathbb{V}[R] \simeq \mathbb{V}[\theta - rp^\star]. \tag{6}$$

[5] We emphasize that this structure allows us to preserve a number of desirable properties common in CARA-Gaussian models. First, the posterior Gaussian, and its mean is a weighted average of the prior mean and average signals. Second, the posterior precision increases linearly with the aggregate precision of signals.

Turning to trading activity, the volume of stocks traded in equilibrium is $T = \|x_B(p^\star)\|$. Since we may normalize the number of outstanding shares to one without loss of generality, T also captures the stock's "turnover" which we will measure in the data.[6] Note that equilibrium prices depend on the realization of signals $\boldsymbol{s}$, so that turnover itself is a random variable. We obtain predictions about the average turnover $\mathbb{E}[T]$.

Our key predictions about return volatility and turnover are as follows:

Proposition 1. *The volatility of returns and average turnover between dates* 1 *and* 2 *satisfy*

$$\frac{\theta_0^2}{r^2}\mathbb{V}[R] \simeq \frac{1}{\rho_R} + q^2\mathbb{V}\left[\theta_B - \theta_R\right], \tag{7}$$

$$\mathbb{E}[T] = \frac{1}{\gamma}\left[\frac{1}{\lambda\rho_B} + \frac{1}{(1-\lambda)\rho_R}\right]^{-1}\mathbb{E}\left[\|\theta_B - \theta_R\|\right]. \tag{8}$$

Social media buzz K increases subsequent return volatility and turnover.
News media buzz decreases subsequent volatility and turnover, as long as the signal-to-noise ratio of news media, $N\rho_\epsilon/\rho_0$, is sufficiently large.

Equations (7) and (8) decompose return volatility and average turnover in equilibrium. The first term in (7) measures the volatility of returns in a fully rational market. This term is equal to the posterior variance which rational investors attach to asset values after observing signals, and therefore captures the *informativeness* of signals. The second term captures the additional volatility due to behavioral biases. The key statistics to determine this second effect are the variance of the *disagreement* between rational and behavioral investors, $\mathbb{V}\left[\theta_B - \theta_R\right]$, and the *confidence* of behavioral investors, as captured by q.

Turnover is also driven by confidence and disagreement effects. The first factor in (8) captures confidence; when posterior beliefs are tighter (high ρ_B and ρ_R) or investors are risk-tolerant (low γ), they trade more aggressively. The second factor is the expected disagreement between rational and behavioral investors, $\mathbb{E}\left[\|\theta_B - \theta_R\|\right]$. As in Harris and Raviv (1993), "differences of opinion make a horse race", and markets are more active when investors disagree about the average quality of assets.

Our central predictions in Proposition 1 are that (i) social media buzz *increases* future volatility and trading activity, and that (ii) news media buzz *decreases* future volatility and trading activity as long as news media are sufficiently informative. The intuition for the impact of social media is simple: Social media buzz does not contribute genuine information. It does, however, boost the confidence of behavioral investors, and also the extent to which they disagree with rational investors. Therefore, social media buzz increases volatility and turnover. The intuition for the impact of news media buzz is more nuanced: On one hand, more information from news media increases investor confidence, thus raising volatil-

[6] Lo and Wang (2000) provide more general formal definitions of volume and turnover.

ity and trading activity. On the other hand, traders become better-informed, and extreme realizations of their beliefs – which lead to disagreement – become less likely. These effects – informativeness and reduced disagreement – dampen volatility and turnover, and they dominate the confidence effect as long as news media are sufficiently informative.

It is important to note that our model is not at odds with the common intuition that news signals increase volatility because prices are sensitive to information. The difference between this argument and our predictions lies in the timing of events: Consider a dynamic version of our model where the stock is also traded at a previous date 0, before investors observe any signals, implying a constant initial market price p_0. If more informative signals arrive at date 1, then the volatility of *past* returns, between date 0 and date 1, is $\mathbb{V}[p^\star/p_0]$, and therefore proportional to the variance of market prices $\mathbb{V}[p^\star]$. It is easy to verify that this *backward-looking* measure of volatility is increasing in the number of informative signals, precisely because prices react strongly to information. However, the volatility of *subsequent* returns, between dates 1 and 2, is $\mathbb{V}[\theta/p^\star]$. This forward-looking measure is characterized in Proposition 1, and it is decreasing in the number of informative signals under the stated conditions. In this paper, we focus on the latter (forward-looking) prediction.

To verify the robustness of our predictions about volatility, Figure 2 presents the results of several simulations of the model, where $\mathbb{V}[R]$ is computed exactly. We note that our analytical approximation in 7 fits very well. Confirming the analytical predictions of Proposition 1, social media buzz always increases volatility, while news media buzz decreases volatility, except in the case where news media is uninformative (low ρ_ϵ and low N).

3.3 Discussion of mechanisms

Our baseline setup models news media as informative signals, and social media as repeated information which is mistaken for news by behavioral traders. We now consider three alternative mechanisms, and note that they produce distinct empirical predictions from the baseline case in Proposition 1. The analytical details are in the online Appendix.

Social media posts as private information

Another alternative is that social media conveys genuine information, but that its signals are only privately accessible to investors who are informed or "well-connected". To see the empirical implications of this setting, we analyze a model where news media is represented by public signals and social media by private ones (see Appendix B.1). In this model, there are no behavioral biases. Following Grossman and Stiglitz (1980), we assume that there is a population of "noise traders", so that the content of private social media signals is not fully revealed by market prices.

Proposition 2. *If social media signals are private information, then the effects of buzz are as follows, as long as there is a sufficient mass of noise traders:*

- *Social and news media buzz both decrease subsequent subsequent return volatility;*
- *When there are many informed traders, news media buzz decreases average turnover and social media buzz increases it;*
- *When there are many uninformed traders, news media buzz increases average turnover and social media buzz decreases it.*

These predictions are intuitive. In a model with private information, volatility is again driven by the informativeness of signals (both private and public), the confidence of both informed and uninformed traders, and the variance of their disagreement. When noise traders are prominent, an increase in the number of signals (private or public) reduces volatility:[7] Confident investors exploit more aggressively the arbitrage opportunities created by noise orders, which prevents large deviations of prices from fundamental values. In terms of trading activity, informed traders place larger orders when their informational advantage over uninformed traders is significant. This occurs when social media buzz is high or when news media buzz is low. When informed traders are prevalent, social media buzz therefore increases turnover while news media buzz decreases it. These predictions are reversed when uninformed traders are prevalent, because they shy away from large orders when find themselves at a significant informational disadvantage.

A limiting case is the Bayesian benchmark discussed in the introduction: When social media conveys information and all agents are informed, it is straightforward to show that both social and news media buzz decrease subsequent volatility and increase turnover. In this case, there is no disagreement among investors. When more signals arrive, everybody's confidence increases, which leads to more trading with noise traders and smaller deviations of prices from fundamental values. Thus in the limiting case, increases in social and news media buzz yield the same qualitative predictions for market activity.

A catalog of behavioral biases

The interpretation of social media may be subject to behavioral biases other than the "echo chamber" effect. We analyze a suite of models where social media produces informative signals, but where its interpretation by some traders deviates from rational behavior.

An advantage of our linear-Gaussian setup is that we can adapt it to allow for most biases which are common in the behavioral finance literature (Hirshleifer, 2001; Barberis

[7] Away from the "noisy" case, the model is not tractable enough to yield clear testable predictions. The analysis in the appendix shows, however, that similar predictions hold true in the "noise-free" case where the mass of noise traders is small, and in the case where the mass of informed traders is sufficiently large. Therefore, the predictions are not an artefact of the "noisy" case.

and Thaler, 2003). Here, we present results from four models of bias which we develop in Appendix B.2. Namely, we consider *overconfident* investors who overstate the precision of their signals as in Scheinkman and Xiong (2003);[8] *conservative* investors who overstate the precision of their prior beliefs as in Barberis et al. (1998);[9] investors who process only a subset of signals due to *rational inattention* as in Sims (2003);[10] and investors with *confirmation bias*, who ignore signals that do not conform with their prior sentiment as in Rabin and Schrag (1999). Proposition 3 summarizes our assumptions and testable predictions in each case.

Proposition 3. *Suppose that there are K informative social media signals with precision ρ_ϵ. The effect of social media buzz with alternative behavioral biases is as follows:*

- ***Overconfidence:*** *Behavioral investors perceive the precision to be $(1+a)\,\rho_\epsilon$.*
  Prediction: If the bias a is small, then social media buzz decreases return volatility and increases turnover. If the bias is large, then social media buzz decreases both volatility and turnover as long as the informativeness of social media $K\rho_\epsilon/\rho_0$ is sufficiently large.

- ***Conservatism:*** *Behavioral investors attach precision $(1+b)\,\rho_0$ to their prior.*
  Prediction: Social media buzz decreases return volatility and increases turnover.

- ***Rational inattention:*** *Behavioral investors observe signals with cognitive noise, which must satisfy a constraint on entropy reduction.*
  Prediction: Social media buzz decreases return volatility and increases turnover.

- ***Confirmation bias:*** *Behavioral traders with an optimistic predisposition process negative signals $s_i < \theta_0$ as if they were equal to the prior mean θ_0.*
  Prediction: Social media buzz increases turnover, and decreases return volatility as long as the number of signals K is sufficiently large.

We note that in each of these models, our decompositions of return volatility and trading activity in (7) and (8) remain valid: Volatility and turnover are determined by the competing effects of informativeness, confidence and disagreement.

Proposition 3 shows that, for most biases, an increase in the number of social media signals tends to reduce volatility and increase turnover. Volatility is reduced because social

[8] Overconfidence is also related to *base rate under-weighting*, where individuals pay too little attention to prior probabilities when updating beliefs (Kahneman and Tversky, 1973). Early experimental evidence was presented by Fischhoff et al. (1977) and Alpert and Raiffa (1982), for example. Odean (1998) reviews the large empirical literature demonstrating overconfidence. Related formal models of overconfidence are in Kyle and Wang (1997), Odean (1998), Daniel et al. (1998) and Gervais and Odean (2001).

[9] Early evidence of conservatism is provided by Edwards (1968). Conservatism is common when individuals have information from large samples, but continue to place higher weight on their prior beliefs than Bayes' rule would suggest (Griffin and Tversky, 1992).

[10] Peng and Xiong (2006) and Kacperczyk et al. (2016), among other papers, also apply rational inattention to the CARA-Gaussian framework to obtain predictions about asset prices and mutual fund returns. By contrast, we focus on predictions about return volatility and trading activity.

media buzz is truly informative in this class of models – unlike in the baseline model of "echo chambers" – and its informativeness leads to reduced volatility. Turnover, however, is increased in most cases because social media signals increases the confidence of investors who bet on perceived disagreements.

Endogenous buzz

Finally, the direction of causation between buzz and market activity could be reversed. Casual empiricism suggests that people become most active on social networks when they disagree with each other.[11] Moreover, recent evidence links the demand for (online and offline) news to uncertainty and market volatility.[12] We analyze these mechanisms formally in Appendix B.3. Proposition 4 summarizes our assumptions and testable predictions in two cases.

Proposition 4. *Suppose that news media buzz and social media buzz are endogenously determined. We consider two models:*

- ***Disagreement shocks:*** *News media are processed rationally, but the posterior mean of behavioral traders is shifted by a random amount with standard deviation δ. Social media buzz is assumed to be an increasing function of disagreement (among other factors).*
  Prediction: *An exogenous increase in δ increases social media buzz and subsequent volatility and turnover. However, conditional on disagreement δ, subsequent volatility and turnover are independent of buzz.*

- ***Uncertainty shocks:*** *The demand for information is increasing in prior uncertainty ρ_0^{-1} (among other factors), and both social and news media buzz respond are produced according to a downward-sloping supply curve.*
  Prediction: *An exogenous increase in uncertainty ρ_0^{-1} increases both news and social media buzz, and it increases volatility as long as the equilibrium response of information is not too strong (i.e. its supply is elastic and its demand is inelastic).*

Proposition 4 shows that increased disagreement can drive up social media activity, return volatility and turnover. However, the predictive power of social media buzz for subsequent market activity should be diminished once we condition on measures of disagreement. If uncertainty shocks raise the demand for information, then we would expect both social media and news media buzz to increase in response. Thus, the correlations between buzz on

[11] See, for example, https://xkcd.com/386/

[12] Kacperczyk et al. (2014; 2016) find support for a model where investors rationally allocate attention to the most uncertain risk factors. Andrei and Hasler (2014) show that investors' attention, as measured by Google searches, is higher in times of high volatility.

social media and subsequent market activity should have the same sign as the correlation between news media buzz and market activity. We now turn to our empirical analysis.

4 Data

This section describes the data we use to measure stock coverage in social and news media, as well as financial data we use to measure stock prices, volatility, trading activity and stock characteristics.

4.1 Measuring 'buzz' in social and news media

We use the Thompson Reuters MarketPsych Index (TRMI) database, which extracts measures of buzz (defined below) and sentiment from English-language news and social media content using a machine learning lexical analysis algorithm. The social and news media sources covered by the index have evolved over time. For consistency, we focus on the period between January 2009 and December 2014.

During this time, the main sources of traditional news media content are (i) Reuters News, (ii) a host of mainstream news sources collected by MarketPsych Data, and (iii) online content collected by Moreover Technologies from about 50,000 internet news sites that include top international and business news sources, top regional new sources, and leading industry sources. The online news content includes many finance-specific sites such as Forbes and SeekingAlpha.

The main source of social media content are (i) content collected by MarketPsych Data from internet forums and finance-specific tweets, and (ii) a social media feed constructed by Moreover Technologies, which captures the top 30% of social media content, as ranked by popularity using incoming links, collected from around 4 million social media sources such as chat rooms (including stock-market specific chats), Facebook posts, blogs, micro-blogs and tweets.

From these sources, the TRMI algorithm extracts high-frequency measures of media coverage, sentiment, and events surrounding each of about 3000 US stocks. The TRMI indicators update every minute. We use a dataset which reports them at the daily frequency.

The *total buzz* of a stock on a given day counts the number of words and phrases referring to the stock in the above sources. This number is obtained by first identifying articles and social media posts about a specific stock, and then counting the total number of phrases and words referring to sentiments (for instance, fear, joy or trust) and/or events (for instance, litigation, mergers or layoffs) related to this stock. Therefore, total buzz captures not only which stocks are being mentioned, but also the intensity of discussion of a particular stock, as captured by the quantity of sentiment/event phrases and words. This measure is more informative than the length of the article, since meaningless words (for instance, 'the', 'are', 'in', etc.) are not included. For our main analysis, we use a monthly measure of total buzz, which is obtained by summing across days.

The *relative buzz* of a stock is then defined as the total buzz of a stock on a given month, divided by the total buzz of all stocks mentioned in that month. This calculation is done separately for social and news media content, yielding our key measures of coverage: Social media relative buzz (BuzzS) and news media relative buzz (BuzzN), both of which are continuous variables between zero and one.[13]

We focus on stocks which are traded on NYSE, AMEX and NASDAQ. We follow the literature in excluding regulated utilities (SIC codes 4910-4949), depository institutions (SIC 6000-6099) and holding and investment companies (SIC 6700-6799). This yields a panel of 2613 stocks observed between January 2009 and December 2014, with 162,268 stock-month observations in total.[14]

4.2 Financial data

We merge our measures of buzz with monthly financial data from the Center for Research in Securities Prices (CRSP) and the Compustat database. The main variables of interest are trading activity and the realized idiosyncratic volatility of each stock. Our measure of trading activity is turnover ("Turn"), which is taken directly from the CRSP data.

Our parametric measure of realized idiosyncratic volatility ("iVolp") is constructed in two steps. First, for every month m in the sample, we estimate a three-factor model of daily returns on each stock by fitting the following regression equation:

$$(R_{it} - Rf_t) = \beta_0^{(m)} + \beta_1^{(m)}(Rm_t - Rf_t) + \beta_2^{(m)} SMB_t + \beta_3^{(m)} HML_t + \epsilon_t^{(m)}.$$

R_{it} is the return to stock i on day t; Rf_t is the one-month treasury bill rate; Rm_t is the return to the value-weighted market portfolio; SMB_t is the average return on the three Fama and French (1993) small-cap portfolios minus the average return on the three big-cap portfolios; and HML_t is the average return on the two value stock portfolios minus the average return on the two growth stock portfolios. Second, we define the idiosyncratic volatility of stock i in month m as the sum of squared errors from this monthly regression. We will check the robustness of our results by considering an alternative, non-parametric,

[13] In our empirical analysis, we focus on *relative buzz*. In the models of Section 3, we interpret increased buzz as an increase in the *total* number of signals about the stock in question. However, the predictions of this model are relevant for our empirical measure because, holding constant the amount of buzz around other assets, an increase in the number of signals about an individual stock corresponds to an increase in relative buzz also. Moreover, we show in Section 5.4 that the predictions of the echo chamber model for volatility also hold true at the market level when total buzz is considered.

[14] The panel is unbalanced due to the entry and exit of stocks. We show that our main results are preserved when we restrict attention to a balanced panel.

measure of idiosyncratic volatility ("iVoln"), which is obtained by taking the variance of returns of each stock at the monthly frequency.

Our analysis also includes a set of financial variables which have predictive power for volatility and trading activity. We obtain measures of firm size (Size), monthly stock price returns (Ret), and the standard deviation of the last 60 monthly returns (TotalSD) from CRSP. Using the Compustat data, we calculate each firm's leverage (Leverage), and its degree of focus as measured by the Herfindahl-Hirschman index of segment revenue (HHI). We include the share of institutional ownership (InstOwn) from the Thomson Reuters Stock Ownership Summary.[15] We further obtain the dispersion of analyst opinions (AnalystDisp) from the I/B/E/S summary files.[16]

4.3 Summary statistics

Table 1 shows sample means and standard deviations for buzz, volatility and turnover, both for the total stocks and disaggregated by industry. Buzz is measured in percentage points. We winsorize all financial data and the buzz measures at 1% to ensure that our estimates are not driven by outliers.

The average of relative buzz is about 0.02%. Values of buzz range from zero, for stocks which are not mentioned in a given month, to 0.5% for the most talked-about stocks.[17] The standard deviation of buzz is slightly higher for news media than for social media. There is evidence of heterogeneity across industries: Buzz is relatively high in the trade, services, manufacturing and finance industries, and relatively low in agriculture and mining/construction.

Table 2 reports sample averages and standard deviations for our financial control variables, and Table 3 shows contemporaneous correlations, with standard errors in parentheses. Social and news media buzz are strongly correlated. Buzz correlates with firm size, and this correlation is stronger for news than for social media. There is also a strong correlation between buzz and turnover, especially for social media. The contemporaneous correlation between volatility and buzz is positive for social media, and negative (but close to zero) for news media. Table 4 decomposes variation in our panel data into between and within stock variation. For both news and social media, the majority of variation occurs is between stocks.

[15] The literature on stock market volatility demonstrates the predictive power of size (Cheung and Ng, 1992), returns (Duffee, 1995), institutional ownership (Dennis and Strickland, 2002) and trading volume (Schwert, 1989). Trading activity is commonly associated with the absolute value of returns (Karpoff, 1987; Schwert, 1989), institutional ownership (Tkac, 1999) and size (Tkac, 1999; Lo and Wang, 2000).

[16] Dispersion is calculated monthly as the natural log of one plus the standard deviation of analyst forecasts, normalized by the absolute value of mean forecast in a given month.

[17] Before we winsorize the top 1%, we observe stocks which have relative buzz of about 10% on individual days.

To further illustrate the dynamics of news and social media buzz variables in our sample, Figure 1 shows 30-day moving averages of the *total* news and social media buzz of all stocks in our sample. At the market level, both types of buzz go through noticeable swings, and they are positively but not perfectly correlated. Figure 3 illustrates our measure of buzz at the industry level. Panel A plots 30-day moving averages of buzz for each industry, and Panel B shows a comparison of the total quantities of news and social buzz for each industry over the whole sample. This reveals considerable variation in buzz across time and industries. This further translates to variation at the stock level, which we exploit in our analysis of volatility and turnover below.

5 Empirical results

In Section 3, we developed predictions about the relationship between buzz in news and social media, and subsequent return volatility and trading activity.

from several models on the relationship between buzz in social and news media, and subsequent volatility and trading activity. We now take these implications to the data by examining whether buzz indeed predicts volatility and turnover. We begin by reporting our estimates, and Section 6 below relates these findings to our theoretical predictions.

5.1 Main results: Predicting volatility and trading activity

Let i index stocks and t index months. We run a panel regression of each stock i's volatility next month $(t+1)$ on this month's buzz in social and news media, $BuzzS_{i,t}$ and $BuzzN_{i,t}$, stock-level control variables $X_{i,t}$, and this month's volatility $iVolp_{i,t}$. We control for stock and time fixed effects α_i and μ_t:

$$iVolp_{i,t+1} = \alpha_i + \mu_t + \beta_S \times BuzzS_{i,t} + \beta_N \times BuzzN_{i,t} + \gamma \cdot X_{i,t} + \delta \times iVolp_{i,t} + \epsilon_{i,t+1}. \quad (9)$$

Table 5 reports the results from estimating Equation (9). Standard errors are clustered at the stock level and the corresponding t-statistics are in parentheses.

We find that high news media buzz predicts lower future volatility and this is statistically significant at the 1% level. Social media buzz predicts higher future volatility, and this relationship is also statistically significant once we control for stock characteristics in column (3). The sign of the coefficient on social media buzz changes once we add the controls, suggesting that the estimates in columns (1) and (2) were biased by omitted variables. Including time fixed effects in column (4) makes little difference, suggesting that the estimated effects of buzz are not driven by unobserved time trends.

A potential concern is that our estimated coefficients, with opposite signs on news and social media, are driven by the strong positive contemporaneous correlation between news and social media buzz. To check whether this is the case, we introduce the two measures separately in columns (5) and (6). The coefficients have the same sign, magnitude and significance as in columns (3) and (4), which suggests that contemporaneous correlation is not driving our results.

To interpret the economic significance of buzz, suppose that a stock goes from having no buzz in social media to being one of the most talked-about stocks, with a relative buzz of 0.5%. Then, according to the most general specification (column (4)), our measure of the stock's subsequent idiosyncratic volatility rises by about 0.0055 on average. This increase corresponds to about half the average volatility, or just under a third of a standard deviation. For an equivalent change in news media buzz, the stock's subsequent volatility falls by 0.008, about a three quarters of the average volatility, or 40% of a standard deviation.

For trading activity, as measured by turnover, we run the analogous panel regression:

$$Turn_{i,t+1} = \alpha_i + \mu_t + \beta_S \times BuzzS_{i,t} + \beta_N \times BuzzN_{i,t} + \gamma \cdot X_{i,t} + \delta \times Turn_{i,t} + \epsilon_{i,t+1}. \quad (10)$$

Table 6 reports our estimates. Again, standard errors are clustered at the stock level and we include stock fixed effects throughout.

We find that high social media buzz predicts high turnover, and that high news media buzz predicts low turnover. This effect is statistically significant at the 1% level, and does not change much when we include control variables in column (3), or time fixed effects in columns (2) and (4). Introducing social and news media buzz separately in columns (5) and (6) affects our estimates only marginally, suggesting that they are not driven by correlation between social and news media.

To interpret the effects, suppose again that relative social media buzz around a stock rises from zero to the level of the most talked-about stocks at 0.5%. Then according to the most general specification in column (4), turnover increases by about $\frac{1}{2}$, which corresponds to 24% of average turnover, or about a quarter of a standard deviation. For an equivalent rise in news media buzz, turnover falls by about 0.88, which is 43% of average turnover, or 47% of a standard deviation.

Our main results are robust in a variety of alternative specifications. Table 7 reports our baseline panel regressions (9) and (10), controlling for a measure of "sentiment" in news and social media (SentN and SentS) from the TRMI database.[18] Absolute sentiments and negative sentiments (SentN_Neg and SentS_Neg) serve as a proxy for potential unobserved stock-specific events. Only news media sentiments appear to have a significant effect on turnover. Moreover, the estimated effects of news and social buzz change very little from

[18]Sentiment on news and social media (SentN and SentS) is calculated as the difference between the number of "positive" and "negative" references to a stock, as identified by a proprietary machine learning algorithm, and taking into account all words and phrases which count towards the stock's total buzz. To account for asymmetric effects, we also include the number of negative references in isolation (SentN_neg and SentS_neg).

our baseline specification.

Tables 8 and 9 show the results of estimating Equations (9) and (10) on a balanced panel, and Table 10 shows the results of running the regression (9) using the non-parametric measure of volatility. Again, the estimated effects of buzz are quantitatively similar to the baseline model, and significant at the 1% level.

5.2 The role of disagreement

Tables 11 and 12 repeat our panel regressions while controlling for the dispersion of analyst opinions, a common measure of disagreement among investors. The effects of BuzzN and BuzzS are much the same as in the baseline regressions.

Moreover, we check whether the dispersion of analyst opinion has predictive power for future news or social media buzz, as one would expect if disagreement were a major driver of media activity. Table 13 shows the results of panel regressions with news and social media buzz ("BuzzN" and "BuzzS") of month $t+1$ as the dependent variable, analyst dispersion in month t as the independent variable, and a set of control variables. The sign on AnalystDisp is positive in news buzz regressions, but negative in social buzz regressions. These coefficients become insignificant once month fixed effects are included.

5.3 Vector autoregression

To allow for richer dynamic interactions, we run a panel vector autoregression (VAR) including news and social media buzz, return volatility and turnover as endogenous variables. We include two lags of endogenous variables, a set of exogenous control variables, and month and time fixed effects.[19] Table 14 shows the results and Figure 4 illustrates the estimated effects with impulse response functions.

The estimated effects are similar to our baseline panel regressions: An increase in social media buzz leads to a significant increase in volatility and turnover which declines over time. An increase in news media buzz leads to a significant decrease in volatility. We were unable to estimate the impact of news media buzz on turnover with a great deal of precision, as witnessed by the unusually wide confidence intervals. The point estimate of this effect is initially negative but not significant at the 5% level.

[19] To control for fixed effects in a computationally feasible manner, we time-demean the endogenous variables to account for time fixed effects, and apply a Helmert transformations to create forward mean differenced forms which remove stock fixed effects. For a vector of endogenous variables $\hat{\boldsymbol{z}}_{i,t}$, in a panel of time periods $t = 1, ..., T$, the Helmert-transformed endogenous variables are

$$\mathbf{z}_{i,t} = \sqrt{\frac{T-t}{T-t+1}}(\hat{\mathbf{z}}_{it} - \frac{1}{T-t}\sum_{n=t+1}^{T}\hat{\mathbf{z}}_{in}).$$

Based on our estimates, we test for Granger causality among the endogenous variables. The null hypothesis is that for two endogenous variables i and j, a contemporaneous increase in i does not predict a significant subsequent change in j. We find significant effects for all pairs (i, j) at the 1% level, with two exceptions: A shock to social media buzz does not predict a significant change in news media buzz (p-value 0.054), and the impact of turnover on social media buzz is only significant at the 5% level (p-value 0.039).

5.4 Market-level effects

Our analysis so far has focused on stock-level news and social media buzz, stock-level trading activity and idiosyncratic volatility. We now examine whether the same effects are present at the market level, i.e. whether the *total buzz* surrounding all stocks, in either social or news media, has predictive power for aggregate volatility and trading activity.

To study the effect of buzz on market return volatility, we obtain a daily time series of market return volatility from a generalized autoregressive conditional heteroskedasticity (ARCH) models. We use the GJR-GARCH model to capture the potential leverage effect, i.e. the asymmetry in the effect of positive and negative return on volatility (see, for example, Duffee, 1995).

We use two alternative series for market return: the value-weighted return from CRSP (VWRet), and the return on the S&P 500 (SPRet). We construct market-level measures of buzz in news media (Mkt_BuzzN) and social media (Mkt_BuzzS) by summing up the total buzz for all individual stocks on each day. For Mkt_BuzzN and Mkt_BuzzS, we sum up buzz for stocks in our whole sample. For analyzing S&P500 returns, we generate measures of buzz (SP_BuzzN and SP_BuzzS) which aggregate total buzz only for S&P500 stocks.

We report the results in Table 15. The negative effect of news buzz and the positive effect of social buzz on volatility can also be found at the market level, using both all stocks or only S&P 500 stocks. Our analysis of turnover, by contrast, did not yield significant results at the market level.

6 Discussion

We have found that "buzz" in social media consistently predicts high idiosyncratic volatility and high trading activity, as measured by turnover. Moreover, "buzz" in news media consistently predicts low volatility and low trading activity. Since both effect go in opposite directions, we can reject the hypothesis that the processing of information from social and news media is described by the same underlying model. We can further interpret our results by comparing them to the predictions in Section 3.

Our model of social media as an "echo chamber" is able to rationalize our findings: We fail to reject any of the predictions of Proposition 1. Thus, the evidence is consistent with a model in which information on social media is frequently repeated, but some boundedly rational investors treat repeated information as if it were new. In this model, social media activity increases volatility and trading activity by boosting disagreement and confidence among traders. News media can have the opposite effect because it reduces disagreement, which dampens volatility and turnover as long as news are sufficiently informative.

Our panel VAR estimation provides further support for the "echo chamber" hypothesis: An increase in news media activity has a significant positive effect on subsequent social media activity, but conversely, the effect of an increase in social media buzz on news media is not significant. These effects are consistent with the hypothesis that social media repeats existing information.

In Section 3, we considered alternative mechanisms, none of which is fully consistent with the evidence. First, in a model where social media signals are private information, more social media activity dampens volatility by improving the informativeness of markets and reducing the impact of noise traders (Proposition 2). Second, we considered a class of models where social media is informative but processed subject to a variety of behavioral biases. In these models, social media activity also tends to reduce volatility due to its impact on informativeness (Proposition 3). These predictions are contrary to our finding that social media buzz is positively correlated with subsequent volatility.

Finally, since our empirical analysis is not based on exogenous variation in news or social media activity, we must consider the possible endogenous determination of news and social media buzz. Our empirical findings on social media are matched by a model where *disagreement* shocks drive both buzz and market activity (Proposition 4, part 1). However, this model predicts that the effect should be diminished by controlling for measures of disagreement. We fail to find support for this prediction, since our estimated effects are virtually unchanged when we control for disagreement, as measured by the dispersion of analyst forecasts.

Finally, we consider the possibility that movements in *uncertainty* determine both the demand for information and stock market activity (Proposition 4, part 2). This model predicts that social and news media are either both positively correlated with subsequent volatility – if the production of news is insufficient to offset a surge in uncertainty – or both negatively correlated with subsequent volatility. By contrast, the empirical results consistently show that the correlation of news and social media with subsequent volatility have opposite signs.

7 Conclusions

In this paper, we have studied whether information from social media is processed in a different way from information from (online or offline) news media. Using a new dataset on news and social media coverage in financial markets, we have established we have established two robust facts about the relationship between such coverage and market activity. First, intense social media coverage predicts high volatility of returns and high trading activity over the next month. Second, intense news media coverage predicts low volatility and low trading activity. These effects are statistically and economically significant. The difference between the impact of social and news media is striking, and new to the literature.

We have related these results to the predictions of a theory of social media as an "echo chamber" which repeats existing information. Our empirical results are consistent with this model. Moreover, we have used formal models to establish that our empirical results are not consistent with models where social media provides private information to a subset of investors, or where social media is informative but misinterpreted according to common behavioral biases. We have also developed tests and models which suggest that our results are not entirely driven by endogenously determined buzz. Nonetheless, our analysis suggests scope for further research on the different roles of news and social media, and it will be valuable in future work to exploit exogenous variation in social or news media activity.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar, "Bayesian Learning in Social Networks.," *Review of Economic Studies*, 2011, *78* (4), 1201 – 1236.

Alpert, Marc and Howard Raiffa, "A progress report on the training of probability assessors," in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982, pp. 294–305.

Andrei, Daniel and Michael Hasler, "Investor attention and stock market volatility," *Review of Financial Studies*, 2014, *28* (1), 33–72.

Antweiler, Werner and Murray Z Frank, "Is all that talk just noise? The information content of internet stock message boards," *Journal of Finance*, 2004, *59* (3), 1259–1294.

Barber, Brad M and Terrance Odean, "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," *Review of Financial Studies*, 2008, *21* (2), 785–818.

Barberis, Nicholas and Richard Thaler, "A survey of behavioral finance," *Handbook of the Economics of Finance*, 2003, *1*, 1053–1128.

_ , Andrei Shleifer, and Robert Vishny, "A model of investor sentiment," *Journal of Financial Economics*, 1998, *49* (3), 307–343.

Bushee, Brian J, John E Core, Wayne Guay, and Sophia JW Hamm, "The role of the business press as an information intermediary," *Journal of Accounting Research*, 2010, *48* (1), 1–19.

Chamley, Christophe, *Rational Herds*, Cambridge University Press, 2004.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang, "Wisdom of crowds: The value of stock opinions transmitted through social media," *Review of Financial Studies*, 2014, *27* (5), 1367–1403.

Cheung, Yin-Wong and Lilian K. Ng, "Stock price dynamics and firm size: An empirical investigation," *Journal of Finance*, 1992, *47* (5), 1985–1997.

Coval, Joshua D and Tyler Shumway, "Is sound just noise?," *Journal of Finance*, 2001, *56* (5), 1887–1910.

Cutler, David M, James M Poterba, and Lawrence H Summers, "What moves stock prices?," *Journal of Portfolio Management*, 1989, *15* (3), 4–12.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, "The sum of all FEARS: Investor sentiment and asset prices," *Review of Financial Studies*, 2015, *28* (1), 1–32.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, "Investor psychology and security market under-and overreactions," *Journal of Finance*, 1998, *53* (6), 1839–1885.

De Bondt, Werner FM and Richard Thaler, "Does the stock market overreact?," *Journal of Finance*, 1985, *40* (3), 793–805.

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 2003, *118* (3), 909–968.

Dennis, Patrick J. and Deon Strickland, "Who blinks in volatile markets, individuals or institutions?," *The Journal of Finance*, 2002, *57* (5), 1923–1949.

Duffee, Gregory R, "Stock returns and volatility: A firm-level analysis," *Journal of Financial Economics*, 1995, *37* (3), 399–420.

Edwards, Ware, "Conservatism in human information processing," in Benjamin Kleinmutz, ed., *Formal Representation of Human Judgment*, John Wiley and Sons, 1968, pp. 17–52.

Falck, Oliver, Robert Gold, and Stephan Heblich, "E-lections: Voting Behavior and the Internet," *The American Economic Review*, 2014, *104* (7), 2238–2265.

Fama, Eugene F and Kenneth R French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 1993, *33* (1), 3–56.

Fang, Lily and Joel Peress, "Media coverage and the cross-section of stock returns," *Journal of Finance*, 2009, *64* (5), 2023–2052.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein, "Knowing with certainty: The appropriateness of extreme confidence.," *Journal of Experimental Psychology: Human perception and performance*, 1977, *3* (4), 552.

Gentzkow, Matthew, "Valuing new goods in a model with complementarity: Online newspapers," *The American Economic Review*, 2007, *97* (3), 713–744.

_ **and Jesse M Shapiro**, "Ideological Segregation Online and Offline," *Quarterly Journal of Economics*, 2011, *126* (4).

Gervais, Simon and Terrance Odean, "Learning to be overconfident," *Review of Financial studies*, 2001, *14* (1), 1–27.

Griffin, Dale and Amos Tversky, "The weighing of evidence and the determinants of confidence," *Cognitive psychology*, 1992, *24* (3), 411–435.

Grossman, Sanford J. and Joseph E. Stiglitz, "On the impossibility of informationally efficient markets," *American Economic Review*, 1980, *70* (3), 393–408.

Harris, Milton and Artur Raviv, "Differences of opinion make a horse race," *Review of Financial Studies*, 1993, *6* (3), 473–506.

Hawkins, Scott and Stephen Hoch, "Low-involvement learning: Memory without evaluation," *Journal of Consumer Research*, 1992, *19*, 212–225.

Hinkley, D. V., "On the ratio of two correlated normal random variables," *Biometrika*, 1969, *56* (3), 635–639.

Hirshleifer, David, "Investor psychology and asset pricing," *Journal of Finance*, 2001, *56* (4), 1533–1597.

Huberman, Gur and Tomer Regev, "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar," *Journal of Finance*, 2001, *56* (1), 387–396.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, "Time-Varying Fund Manager Skill," *The Journal of Finance*, 2014, *69* (4), 1455–1484.

_ , _ , and _ , "A rational theory of mutual funds' attention allocation," *Econometrica*, 2016, *84* (2), 571–626.

Kahneman, Daniel and Amos Tversky, "On the psychology of prediction.," *Psychological review*, 1973, *80* (4), 237.

Karpoff, Jonathan M, "The relation between price changes and trading volume: A survey," *Journal of Financial and quantitative Analysis*, 1987, *22* (01), 109–126.

Kondor, Péter, "The more we know about the fundamental, the less we agree on the price," *Review of Economic Studies*, 2012, *79* (3), 1175–1207.

Kyle, Albert S and F Albert Wang, "Speculation duopoly with agreement to disagree: Can overconfidence survive the market test?," *Journal of Finance*, 1997, *52* (5), 2073–2090.

Lo, Andrew W and Jiang Wang, "Trading volume: Definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 2000, *13* (2), 257–300.

Milgrom, Paul and Nancy Stokey, "Information, trade and common knowledge," *Journal of Economic Theory*, 1982, *26* (1), 17–27.

Muthen, Bengt, "Moments of the censored and truncated bivariate normal distribution," *British Journal of Mathematical and Statistical Psychology*, 1990, *43* (1), 131–143.

Odean, Terrance, "Volume, volatility, price, and profit when all traders are above average," *Journal of Finance*, 1998, *53* (6), 1887–1934.

Peng, Lin and Wei Xiong, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 2006, *80* (3), 563–602.

Pouget, Sebastien, Julien Sauvagnat, and Stephane Villeneuve, "A Mind is a Terrible Thing to Change: Confirmatory Bias in Financial Markets," 2014. Mimeo.

Rabin, Matthew and Joel L Schrag, "First impressions matter: A model of confirmatory bias," *Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

Reuters Institute for the Study of Journalism, "Digital News Report," Technical Report, 2016.

Scheinkman, José A. and Wei Xiong, "Overconfidence and speculative bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Schwert, G. William, "Why does stock market volatility change over time?," *Journal of Finance*, 1989, *44* (5), 1115–1153.

Sims, Christopher A, "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Tetlock, Paul C, "Giving content to investor sentiment: The role of media in the stock market," *Journal of Finance*, 2007, *62* (3), 1139–1168.

_ , "All the news that's fit to reprint: Do investors react to stale information?," *Review of Financial Studies*, 2011, *24* (5), 1481–1512.

Tkac, Paula A, "A trading volume benchmark: Theory and evidence," *Journal of Financial and Quantitative Analysis*, 1999, *34* (01), 89–114.

Tables and figures

Table 1: Summary statistics for buzz, volatility and turnover

Industries		**BuzzN**	**BuzzS**	**iVolp**	**Turn**
Agriculture	Mean	0.0118	0.0074	0.0108	1.4236
	SD	0.0197	0.0102	0.0179	1.8153
	Max	0.2315	0.0687	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	520	520	520	520
Mining&Constr	Mean	0.0174	0.0264	0.0131	3.0775
	SD	0.0314	0.0688	0.0198	2.5168
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	11106	11106	11106	11106
Manufacturing	Mean	0.0288	0.0314	0.0111	1.9947
	SD	0.0657	0.0897	0.0197	1.8134
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	71569	71569	71569	71569
Utilities	Mean	0.0330	0.0240	0.0107	1.9643
	SD	0.0793	0.0667	0.0202	1.7116
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	9705	9705	9705	9705
Trade	Mean	0.0323	0.0225	0.0093	2.1809
	SD	0.0657	0.0647	0.0164	1.7985
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	17709	17709	17709	17709
Finance	Mean	0.0311	0.0229	0.0073	1.5866
	SD	0.0685	0.0781	0.0156	1.5239
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	13376	13376	13376	13376
Services	Mean	0.0327	0.0304	0.0111	1.9164
	SD	0.0803	0.0965	0.0190	1.7618
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	30908	30908	30908	30908
Other	Mean	0.0239	0.0270	0.0144	1.8484
	SD	0.0664	0.0834	0.0205	1.9401
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	7375	7375	7375	7375
Total	Mean	0.0293	0.0285	0.0109	2.0303
	SD	0.0682	0.0849	0.0191	1.8644
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	162268	162268	162268	162268

Table 2: Summary statistics for control variables

	Size	**InstOwn**	**Ret**	**HHI**	**Leverage**	**TotalSD**
Mean	13.8058	0.6972	0.0196	0.7389	0.0258	0.1430
SD	1.7366	0.2438	0.1274	0.2923	0.0546	0.0632
Max	18.3953	1.0000	0.4822	1.0000	0.3417	0.3791
Min	10.1016	0.0597	-0.3136	0.0000	0.0000	0.0489
N	159358	162268	161526	141532	151185	162220

Table 3: Contemporaneous correlations

Variables	BuzzN	BuzzS	iVolp	Turn	Size	InstOwn	Ret	HHI	Leverage	TotalSD
BuzzN	1.000									
BuzzS	0.554	1.000								
	(0.000)									
iVolp	-0.032	0.100	1.000							
	(0.000)	(0.000)								
Turn	0.133	0.307	0.258	1.000						
	(0.000)	(0.000)	(0.000)							
Size	0.409	0.230	-0.386	0.176	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)						
InstOwn	0.012	-0.085	-0.174	0.235	0.339	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
Ret	-0.001	-0.010	0.120	0.023	-0.042	-0.010	1.000			
	(0.691)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
HHI	-0.076	-0.008	0.076	0.049	-0.173	-0.052	0.003	1.000		
	(0.000)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)	(0.242)			
Leverage	0.059	0.051	0.065	-0.008	-0.027	-0.044	0.007	-0.026	1.000	
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.006)	(0.000)		
TotalSD	-0.132	0.077	0.331	0.185	-0.470	-0.166	0.020	0.109	0.077	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table 4: Summary statistics between and within stocks

Variable		Mean	Std. Dev.	Min	Max	Observations
BuzzN	overall	0.030	0.069	0	0.532	N = 133056
	between		0.063	0	0.532	n = 1848
	within		0.028	-0.258	0.517	T = 72
BuzzS	overall	0.030	0.087	0	0.650	N = 133056
	between		0.075	0	0.650	n = 1848
	within		0.043	-0.452	0.667	T = 72
iVolp	overall	0.010	0.018	0.000	0.131	N = 133056
	between		0.009	0.001	0.066	n = 1848
	within		0.015	-0.052	0.138	T = 72
Turn	overall	2.035	1.832	0.106	11.006	N = 133056
	between		1.397	0.113	10.616	n = 1848
	within		1.186	-4.206	12.273	T = 72

Table 5: Volatility Regressions

	(1) iVolp(t+1)	(2) iVolp(t+1)	(3) iVolp(t+1)	(4) iVolp(t+1)	(5) iVolp(t+1)	(6) iVolp(t+1)
iVolp	0.302***	0.225***	0.213***	0.185***	0.187***	0.183***
	(34.70)	(27.52)	(24.20)	(21.21)	(21.48)	(20.98)
BuzzN	-0.0199***	-0.0187***	-0.0193***	-0.0167***	-0.0138***	
	(-7.86)	(-8.19)	(-7.44)	(-6.85)	(-5.51)	
BuzzS	-0.00524**	0.000559	0.00991***	0.0113***		0.00934***
	(-2.23)	(0.24)	(4.15)	(4.85)		(4.06)
Size			-0.00791***	-0.00618***	-0.00610***	-0.00620***
			(-34.30)	(-21.70)	(-21.28)	(-21.82)
InstOwn			0.00462***	0.00298***	0.00278**	0.00305***
			(4.31)	(2.70)	(2.49)	(2.76)
HHI			0.000211	-0.0000732	-0.0000593	-0.0000969
			(0.31)	(-0.11)	(-0.09)	(-0.15)
Leverage			0.00282	0.00354	0.00354	0.00353
			(1.24)	(1.52)	(1.51)	(1.51)
Turn			-0.000235***	-0.000317***	-0.000206**	-0.000356***
			(-2.88)	(-3.59)	(-2.29)	(-4.02)
Ret			-0.0108***	-0.00952***	-0.00964***	-0.00947***
			(-19.29)	(-14.77)	(-14.91)	(-14.66)
TotalSD			-0.0769***	-0.0627***	-0.0620***	-0.0623***
			(-13.43)	(-9.98)	(-9.80)	(-9.90)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes	Yes	Yes
N	159549	159549	131878	131878	131878	131878
R^2	0.092	0.159	0.142	0.178	0.177	0.177

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Turnover Regressions

	(1) Turn(t+1)	(2) Turn(t+1)	(3) Turn(t+1)	(4) Turn(t+1)	(5) Turn(t+1)	(6) Turn(t+1)
Turn	0.584***	0.590***	0.589***	0.595***	0.604***	0.591***
	(89.66)	(88.03)	(75.51)	(74.12)	(72.34)	(72.78)
BuzzN	-1.841***	-1.974***	-1.670***	-1.765***	-1.519***	
	(-7.21)	(-7.71)	(-6.03)	(-6.38)	(-5.25)	
BuzzS	0.865***	0.852***	1.014***	0.961***		0.748***
	(5.23)	(5.21)	(5.21)	(4.96)		(3.84)
Size			-0.0259	0.0628***	0.0704***	0.0605***
			(-1.64)	(3.22)	(3.57)	(3.13)
InstOwn			0.431***	0.304***	0.288***	0.312***
			(6.34)	(4.44)	(4.18)	(4.53)
HHI			0.0384	0.00204	0.00327	-0.000525
			(0.82)	(0.04)	(0.07)	(-0.01)
Leverage			0.123	0.0559	0.0560	0.0551
			(0.98)	(0.45)	(0.45)	(0.44)
iVolp			-5.655***	-6.107***	-5.932***	-6.281***
			(-14.37)	(-15.46)	(-15.08)	(-15.80)
AbsRet			0.498***	0.395***	0.391***	0.385***
			(9.78)	(7.78)	(7.70)	(7.56)
TotalSD			-0.953***	-0.435	-0.376	-0.392
			(-3.18)	(-1.30)	(-1.10)	(-1.17)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes	Yes	Yes
N	159613	159613	131891	131891	131891	131891
R^2	0.346	0.384	0.340	0.380	0.379	0.379

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Regressions with Sentiment Controls

	(1) iVolp(t+1)	(2) iVolp(t+1)	(3) Turn(t+1)	(4) Turn(t+1)
iVolp	0.179***	0.155***	-6.304***	-6.624***
	(17.84)	(15.66)	(-12.94)	(-13.44)
Turn	-0.0000971	-0.000197**	0.579***	0.583***
	(-1.26)	(-2.37)	(72.05)	(69.20)
BuzzN	-0.0176***	-0.0141***	-1.636***	-1.688***
	(-7.11)	(-6.35)	(-5.97)	(-6.29)
BuzzS	0.00794***	0.00924***	1.099***	1.057***
	(3.67)	(4.29)	(5.38)	(5.13)
SentN	0.00161***	0.00116***	0.190***	0.161***
	(3.88)	(2.84)	(7.01)	(6.14)
SentN_neg	-0.00140*	0.000338	-0.238***	-0.0709
	(-1.66)	(0.41)	(-4.47)	(-1.39)
SentS	0.0000597	0.000106	-0.0290	0.00290
	(0.20)	(0.36)	(-1.41)	(0.14)
SentS_neg	0.000184	0.000463	0.0347	-0.0118
	(0.32)	(0.82)	(0.89)	(-0.31)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	116491	116491	116504	116504
R^2	0.120	0.157	0.331	0.378

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Volatility Regressions with Balanced Panel

	(1) iVolp(t+1)	(2) iVolp(t+1)	(3) iVolp(t+1)	(4) iVolp(t+1)
iVolp	0.338***	0.256***	0.241***	0.212***
	(33.73)	(26.53)	(24.19)	(21.36)
BuzzN	-0.0180***	-0.0175***	-0.0182***	-0.0158***
	(-7.08)	(-7.55)	(-7.46)	(-6.75)
BuzzS	-0.00445**	0.00125	0.00898***	0.0104***
	(-2.01)	(0.58)	(3.72)	(4.34)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	131208	131208	112515	112515
R^2	0.117	0.184	0.163	0.197

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Volatility Regressions with Balanced Panel

	(1) Turn1	(2) Turn1	(3) Turn1	(4) Turn1
Turn	0.609***	0.614***	0.613***	0.619***
	(86.29)	(83.46)	(75.35)	(73.07)
BuzzN	-1.791***	-1.956***	-1.655***	-1.755***
	(-7.67)	(-8.45)	(-6.38)	(-6.71)
BuzzS	0.720***	0.717***	0.920***	0.852***
	(4.25)	(4.28)	(4.71)	(4.36)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	131208	131208	112515	112515
R^2	0.373	0.416	0.371	0.415

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Non-Parametric Volatility Regressions

	(1)	(2)	(3)	(4)
	iVoln1	iVoln1	iVoln1	iVoln1
iVoln	0.407***	0.282***	0.321***	0.247***
	(52.81)	(35.83)	(41.23)	(29.70)
BuzzN	-0.0327***	-0.0281***	-0.0337***	-0.0268***
	(-8.07)	(-8.37)	(-7.81)	(-7.28)
BuzzS	-0.0119***	0.00179	0.0103***	0.0154***
	(-3.64)	(0.55)	(2.90)	(4.52)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	159655	159655	131922	131922
R^2	0.171	0.312	0.213	0.326

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Volatility Regressions

	(1)	(2)	(3)	(4)
	iVolp(t+1)	iVolp(t+1)	iVolp(t+1)	iVolp(t+1)
iVolp	0.302***	0.225***	0.178***	0.154***
	(34.70)	(27.52)	(17.79)	(15.59)
BuzzN	-0.0199***	-0.0187***	-0.0179***	-0.0148***
	(-7.86)	(-8.19)	(-7.19)	(-6.58)
BuzzS	-0.00524**	0.000559	0.00790***	0.00913***
	(-2.23)	(0.24)	(3.65)	(4.24)
AnalystDisp			-0.00144**	-0.00128*
			(-2.11)	(-1.91)
N	159549	159549	116491	116491
R^2	0.092	0.159	0.120	0.157
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Turnover Regressions

	(1) Turn(t+1)	(2) Turn(t+1)	(3) Turn(t+1)	(4) Turn(t+1)
Turn	0.584***	0.590***	0.578***	0.581***
	(89.66)	(88.03)	(72.13)	(69.03)
BuzzN	-1.841***	-1.974***	-1.631***	-1.735***
	(-7.21)	(-7.71)	(-6.01)	(-6.45)
BuzzS	0.865***	0.852***	1.098***	1.053***
	(5.23)	(5.21)	(5.37)	(5.10)
AnalystDisp			0.0336	0.0179
			(0.84)	(0.48)
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	159613	159613	116504	116504
R^2	0.346	0.384	0.330	0.378

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Buzz Regressions

	(1) BuzzN(t+1)	(2) BuzzN(t+1)	(3) BuzzS(t+1)	(4) BuzzS(t+1)
BuzzN	0.250***	0.246***	-0.0854***	-0.0865***
	(5.31)	(5.20)	(-5.99)	(-6.01)
BuzzS	0.0264***	0.0277***	0.678***	0.680***
	(2.58)	(2.71)	(34.00)	(33.89)
AnalystDisp	0.00312***	0.00151	-0.00141	-0.00195
	(2.67)	(1.31)	(-1.10)	(-1.54)
Size	-0.00112**	0.00153***	0.00238***	0.00377***
	(-2.57)	(2.69)	(4.45)	(4.75)
InstOwn	0.00245	0.0000383	-0.00203	-0.00320
	(1.49)	(0.02)	(-0.92)	(-1.33)
HHI	0.00284	0.00178	0.000205	-0.000158
	(1.48)	(0.93)	(0.09)	(-0.07)
Leverage	0.00873*	0.00618	0.00140	0.000473
	(1.89)	(1.36)	(0.33)	(0.11)
Turn	-0.0000575	-0.000279	-0.0000735	-0.000201
	(-0.30)	(-1.37)	(-0.27)	(-0.70)
AbsRet	-0.00114	-0.00243*	-0.00159	-0.00286*
	(-0.82)	(-1.66)	(-0.98)	(-1.69)
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	116531	116531	116531	116531
R^2	0.070	0.077	0.445	0.446

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Panel VAR by Size Quartiles

	(1)	(2)	(3)	(4)	(5)
	All	Q1	Q2	Q3	Q4
iVolp					
BuzzN(t-1)	-0.0132***	-0.0537***	-0.0374***	0.00258	-0.00445**
	(-7.18)	(-5.33)	(-6.02)	(0.59)	(-2.28)
BuzzN(t-2)	-0.00588***	-0.0444***	-0.00896	0.00926**	-0.00237
	(-3.43)	(-4.76)	(-1.25)	(2.17)	(-1.39)
BuzzS(t)	0.0111***	0.00695	0.0198***	0.0143***	0.00505**
	(4.23)	(1.09)	(2.62)	(2.62)	(1.96)
BuzzS(t-2)	0.00130	0.00185	-0.00999*	0.00787	0.00299
	(0.56)	(0.31)	(-1.67)	(1.45)	(1.24)
Turn					
BuzzN(t-1)	-0.219	0.0302	-1.028**	-0.672	0.135
	(-1.29)	(0.06)	(-2.37)	(-1.53)	(0.61)
BuzzN(t-2)	0.583***	0.735	0.484	-0.0758	0.382*
	(3.67)	(1.54)	(0.95)	(-0.18)	(1.94)
BuzzS(t-1)	1.303***	1.405***	1.655***	1.133**	1.027***
	(6.47)	(3.20)	(3.09)	(2.23)	(4.15)
BuzzS(t-2)	-0.0834	-0.830**	-1.144**	0.637	0.754***
	(-0.47)	(-2.19)	(-2.56)	(1.33)	(3.37)
Controls	Yes	Yes	Yes	Yes	Yes
N	127511	31877	31878	31878	31878

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Market-Level GJR-GARCH

	(1) VWRet	(2) VWRet	(3) SPRet	(4) SPRet	(5) SPRet
Main					
Constant	0.000687***	0.000885***	0.000573**	0.000780***	0.000728**
	(2.84)	(3.00)	(2.41)	(2.71)	(2.50)
ARCH					
ARCH(-1)	0.206***	0.271***	0.211***	0.309***	0.310***
	(6.03)	(5.01)	(5.80)	(4.95)	(5.06)
TARCH(-1)	-0.0611**	-0.146***	-0.0756**	-0.179***	-0.207***
	(-1.98)	(-2.59)	(-2.32)	(-2.82)	(-3.36)
GARCH(-1)	1.163***	0.540***	1.165***	0.544***	0.572***
	(12.49)	(8.52)	(11.90)	(8.48)	(11.33)
Constant	-0.0000462***		-0.0000435***		
	(-4.77)		(-4.41)		
HET					
MktBuzzN(-1)		-0.0000185***		-0.0000179***	
		(-4.31)		(-4.16)	
MktBuzzS(-1)		0.00000921***		0.00000876***	
		(3.23)		(3.08)	
SPBuzzN(-1)					-0.0000305***
					(-5.20)
SPBuzzS(-1)					0.0000270***
					(4.61)
Constant		-8.596***		-8.711***	-9.056***
		(-34.40)		(-35.35)	(-34.61)
N	1510	1183	1510	1183	1183
ll	4665.3	3704.0	4693.1	3721.2	3727.4

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figures

Figure 1: **News and social media buzz at the market level**

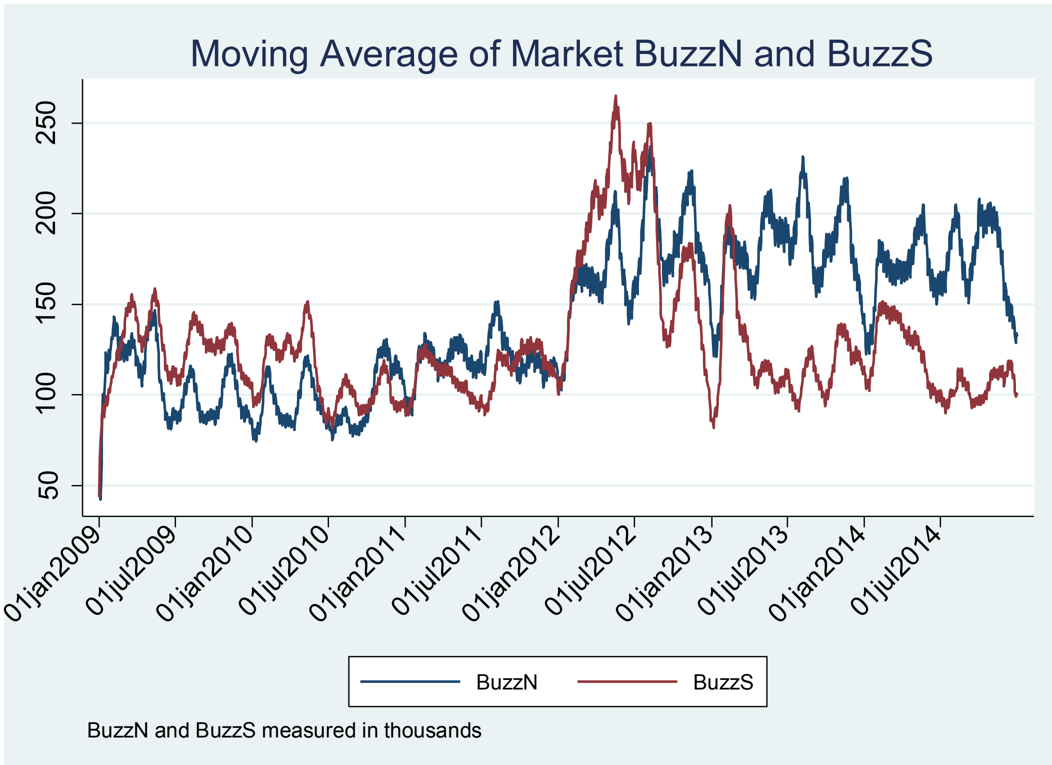


Figure 2: **Simulated comparative statics.** The panels above show the effect of increasing news and social media buzz. The dashed line in each panel is the approximate volatility of returns computed from Equation 7. The solid line is the exact volatility, computed by taking $J = 10000$ draws of θ and $\boldsymbol{s}$, computing equilibrium prices and returns for each draw, and taking the sample variance of returns across all draws. The parameters are $\rho_0 = 2$, $\theta_0 = 20$, $\lambda = 0.5$, $r = 0.1$ and $\gamma = 2$. For comparative statics with respect to N (news media), we set $K = 5$; for comparative statics with respect to K (social media), we set $N = 5$. The high signal-noise ratio case has $\rho_\epsilon = 0.5$, the low signal-noise ratio case has $\rho_\epsilon = 0.1$.

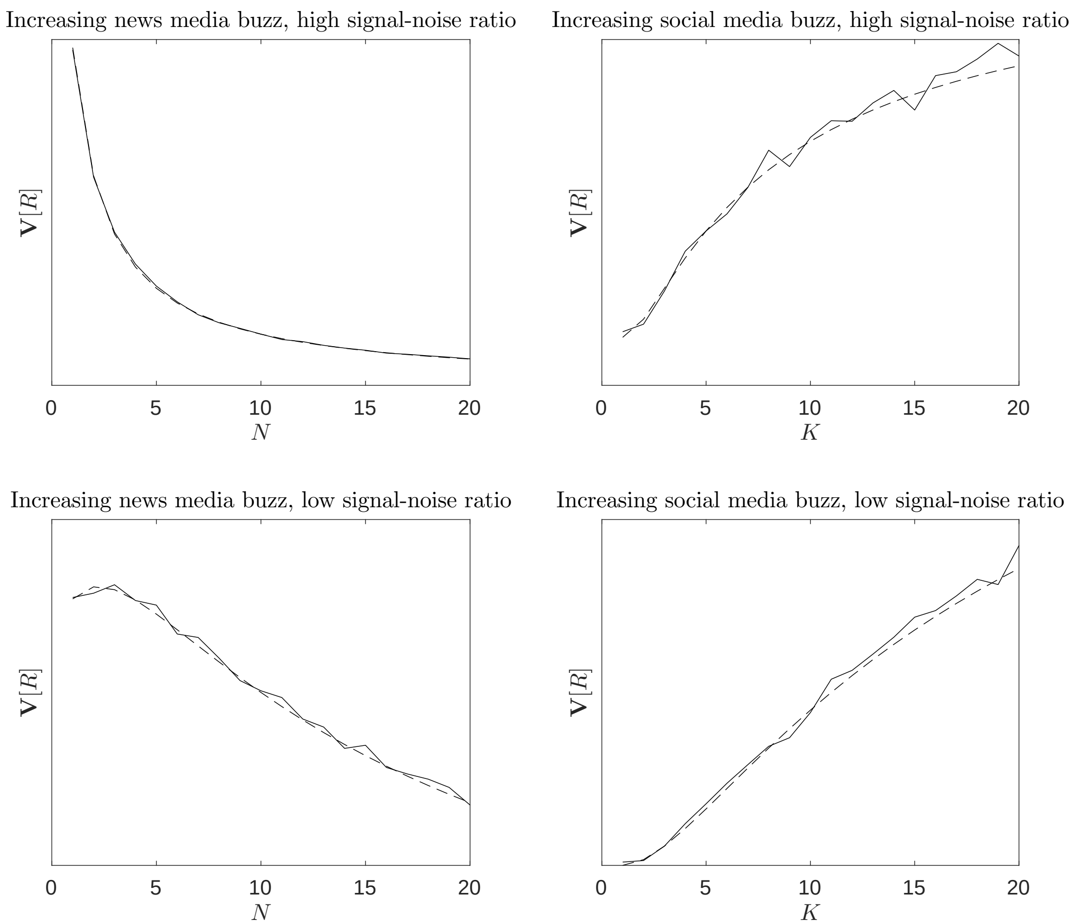
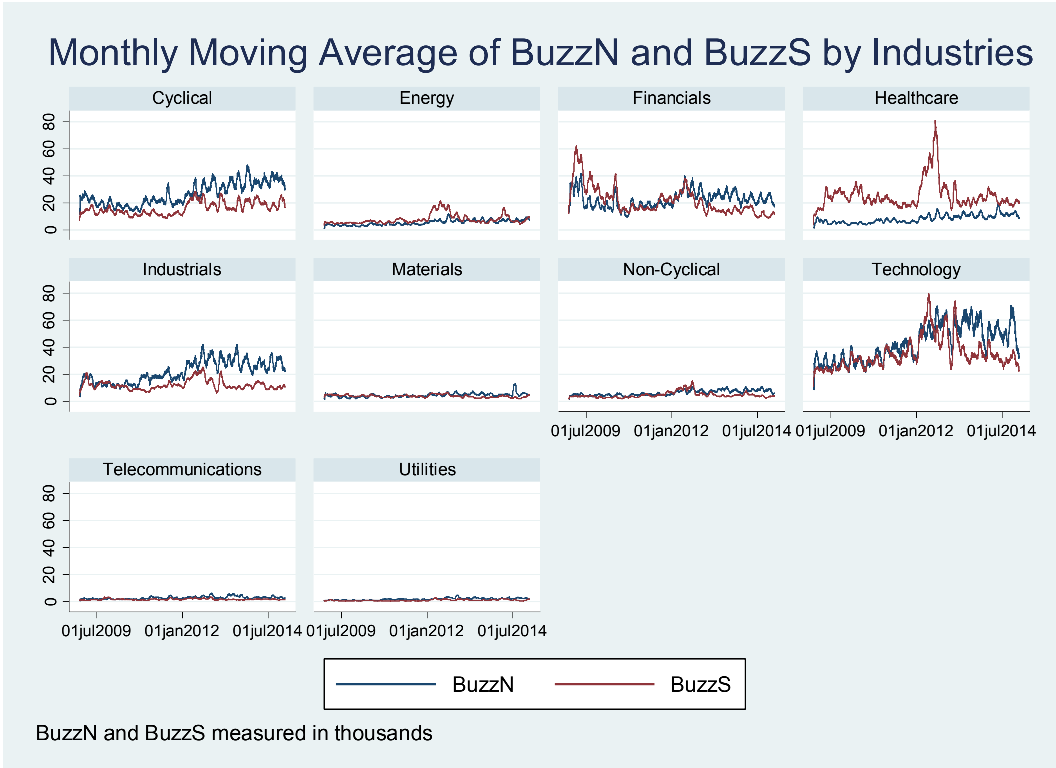
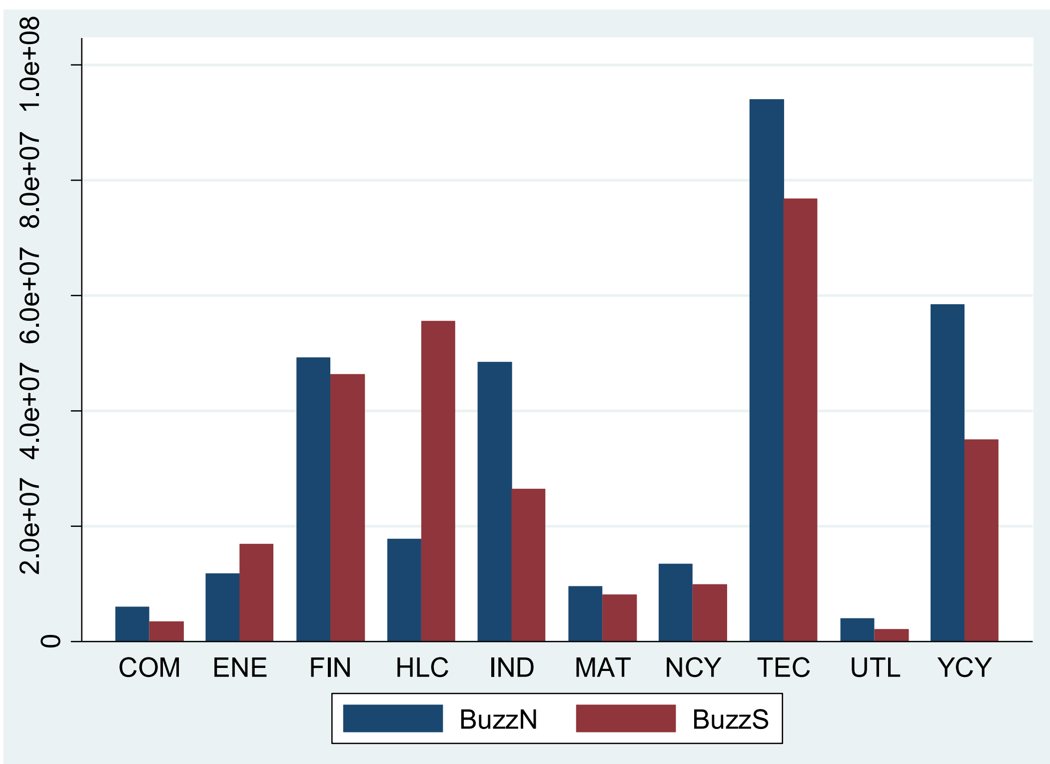


Figure 3: **News and social media buzz by industry**

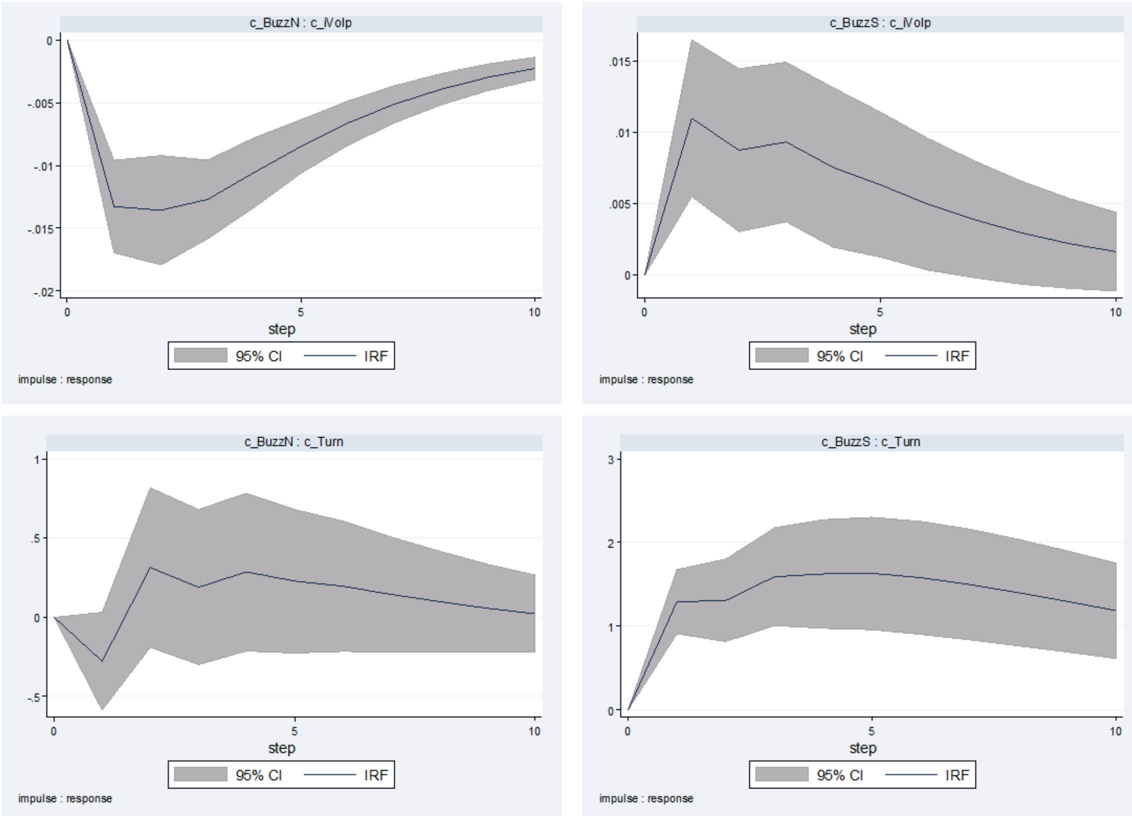


Panel A.



Panel B.

Figure 4: **Impulse response functions**



Appendix for online publication

A Proof of Proposition 1

We begin by characterizing the difference in beliefs, $\theta_B - \theta_R \equiv \Delta$. From (3), and using the definition of repeated signals, we have

$$\begin{aligned}\theta_B &= \sum_{i=1}^{NK} \frac{\rho_\epsilon}{\rho_B} s_i + \left(1 - NK\frac{\rho_\epsilon}{\rho_B}\right)\theta_0 \\ &= \sum_{i=1}^{N} \frac{K\rho_\epsilon}{\rho_B} s_i + \left(1 - N\frac{K\rho_\epsilon}{\rho_B}\right)\theta_0,\end{aligned}$$

and subtracting (1) gives

$$\Delta = \theta_B - \theta_R = N\left(w_B - w_R\right)\left(\bar{s} - \theta_0\right),$$

where $\bar{s} = N^{-1}\sum_{i=1}^{N} s_i$ is the average informative signal, and $w_B = K\rho_\epsilon/\rho_B$ and $w_R = \rho_\epsilon/\rho_R$ denote the weights placed on each informative signal by behavioral and rational traders respectively. Note that Δ is normally distributed with mean zero and variance

$$\begin{aligned}\mathbb{V}\left[\Delta\right] &= N^2\left(w_B - w_R\right)^2\mathbb{V}\left[\bar{s}\right] \\ &= N^2\left(w_B - w_R\right)^2\left(\frac{1}{\rho_0} + \frac{1}{N\rho_\epsilon}\right). \qquad (11)\end{aligned}$$

The last line follows from $\bar{s} = \theta + N^{-1}\sum_{i=1}^{N}\epsilon_i$ and the independence of ϵ_i from θ and each other.

Properties of return volatility

Using the law of total variance in Equation (6), we have

$$\frac{\theta_0^2}{r^2}\mathbb{V}[R] \;\; \simeq \;\; \mathbb{V}[\theta - rp^\star] = \mathbb{E}\left[\mathbb{V}[\theta - rp^\star|\mathbf{s}]\right] + \mathbb{V}\left[\mathbb{E}[\theta - rp^\star|\mathbf{s}]\right].$$

Since the price is known conditional on signals, we have $\mathbb{V}(\theta - rp^\star|\mathbf{s}) = \mathbb{V}(\theta|\mathbf{s}) = \rho_R^{-1}$. Moreover, rational beliefs satisfy $\mathbb{E}\left[\theta|\boldsymbol{s}\right] = \theta_R$, implying $\mathbb{E}[\theta - rp^\star|\mathbf{s}] = q(\theta_R - \theta_B)$. Combining, we obtain Equation (7). Using (11) to evaluate $\mathbb{V}\left[\Delta\right]$, substituting for ρ_B and ρ_R from (2) and (4), and simplifying, we have

$$
\begin{aligned}
\frac{\theta_0^2}{r^2}\mathbb{V}[R] &\simeq \frac{1}{\rho_R} + q^2\mathbb{V}\left[\Delta\right] \\
&= \frac{\rho_0 + N\rho_\epsilon\left[1+\lambda\left(K-1\right)\right]^2}{\left[\rho_0 + \left[1+\lambda(K-1)\right]N\rho_\epsilon\right]^2}.
\end{aligned}
$$

For comparative statics with respect to social media buzz K , consider the monotone transformation $\xi \equiv 1 + \lambda(K-1)$. We have

$$
\frac{\theta_0^2}{r^2}\frac{\partial\mathbb{V}\left[R\right]}{\partial\xi} = \frac{2N\rho_\epsilon\rho_0(\xi-1)}{\left[\rho_0 + \xi N\rho_\epsilon\right]^3} > 0,
$$

so that volatility is increasing in K. For news media buzz N, we have

$$
\frac{\theta_0^2}{r^2}\frac{\partial\mathbb{V}[R]}{\partial N} = \frac{\left[1+\lambda(K-1)\right]\rho_\epsilon}{\left[\rho_0 + \left(1+\lambda(K-1)\right)N\rho_\epsilon\right]^3}\left\{\left[\lambda(K-1)-1\right]\rho_0 - N\rho_\epsilon\left[1+\lambda(K-1)\right]^2\right\},
$$

and it follows that $\frac{\partial\mathbb{V}[R]}{\partial N} < 0$ if and only if

$$
\frac{N\rho_\epsilon}{\rho_0} > \frac{\lambda(K-1)-1}{\left[\lambda(K-1)+1\right]^2},
$$

which is satisfied as long as the signal-noise ratio is sufficiently large, as required.

Properties of turnover

The decomposition in Equation (8) follows directly by evaluating $T = \|x_B(p)\|$, inserting the equilibrium price from (5) and taking expectations on both sides. The absolute value of disagreement $\|\Delta\|$ has a folded normal distribution, and therefore its expectation is

$$
\mathbb{E}\left[\|\Delta\|\right] = \sqrt{\frac{2\mathbb{V}\left[\Delta\right]}{\pi}}.
$$

Using (11), substituting for ρ_B and ρ_R from (2) and (4), and simplifying, we have

$$
\gamma\mathbb{E}\left[T\right] = \Psi \times \frac{(K-1)\sqrt{N\rho_0 + N^2\rho_\epsilon}}{\rho_0 + \left(1+\lambda(K-1)\right)N\rho_\epsilon},
$$

where the constant $\Psi \equiv \sqrt{\frac{2\rho_0\rho_\epsilon}{\pi}}\lambda(1-\lambda)$ is independent of the number of signals N and K. For comparative statics, consider the monotone transformation of $\mathbb{E}\left[T\right]$,

$$
\tau \equiv \left(\frac{\gamma}{\Psi}\mathbb{E}[T]\right)^2 = \frac{(K-1)^2\left[N\rho_0 + N^2\rho_\epsilon\right]}{\left[\rho_0 + \left(1+\lambda(K-1)\right)N\rho_\epsilon\right]^2}.
$$

Differentiating with respect to K,

$$\frac{\partial \tau}{\partial K} = \frac{2N(K-1)\left(\rho_0 + N\rho_\epsilon\right)^2}{\left[\rho_0 + \left(1 + \lambda(K-1)\right) N\rho_\epsilon\right]^3} > 0,$$

which implies $\frac{\partial \mathbb{E}[T]}{\partial K} > 0$. Differentiating with respect to N,

$$\frac{\partial \tau}{\partial N} = \frac{\rho_0 \left(K-1\right)^2}{\left[\rho_0 + \left(1 + \lambda(K-1)\right) N\rho_\epsilon\right]^3} \times \left\{\rho_0 - \left[\lambda\left(K-1\right) - 1\right] N\rho_\epsilon\right\}.$$

Clearly $\frac{\partial \mathbb{E}[T]}{\partial N} < 0$ if and only if $\frac{\partial \tau}{\partial N} < 0$, or equivalently:

$$\frac{N\rho_\epsilon}{\rho_0} > \frac{1}{\lambda\left(K-1\right) - 1}.$$

This is again true as long as the signal-to-noise ratio is sufficiently large, which completes the proof.

B Alternative models

We formally examine three alternative models of the relationship between social media, news media and stock market activity. We consider social media as private information (Proposition 2), alternative behavioral biases in the interpretation of social media (Proposition 3) and the possibility of endogenous buzz (Proposition 4). Our predictions refer to the approximate formula for return volatility in Equation (7), but go through in exact numerical computations of return volatility which are available from the authors on request.

B.1 Social media as private information

Assume that all investors are rational Bayesians, have a prior $\mathcal{N}\left(\theta_0, \rho_0^{-1}\right)$ and observe a Gaussian news media signal $s_N = \theta + \epsilon_N$ with precision ρ_N. A mass λ of "informed" (or "well-connected") investors observe, in addition, a Gaussian social media signal $s_S = \theta + \epsilon_S$ with precision ρ_S. Let $\boldsymbol{s} = (s_N, s_S)$. There is a mass of noise traders with stochastic demand $Q \sim \mathcal{N}\left(0, \rho_Q^{-1}\right)$ which is independent of price. News and social media "buzz", respectively, are measured by the precision of signals ρ_N and ρ_S. This is analogous to our treatment in Section 3, because increasing the precision of signals is equivalent to increasing the number of signals in a rational Gaussian market. We normalize risk aversion to $\gamma = 1$.

Uninformed investors learn from prices. Using standard arguments (e.g. Chamley, 2004, chapter 15), this is equivalent to observing the signal $Y = S + \frac{Q}{\rho_S}$, which is Gaussian with precision $\rho_Y = \rho_S^2\rho_Q/(1 + \rho_S\rho_Q) < \rho_S$. For informed investors, the posterior precision is

$\rho_I = \rho_0 + \rho_N + \rho_S$. For uninformed investors, the posterior precision is $\rho_U = \rho_0 + \rho_N + \rho_Y$. The posterior means of beliefs are

$$\theta_U = \frac{\rho_0}{\rho_U}\theta_0 + \frac{\rho_N}{\rho_U}N + \frac{\rho_Y}{\rho_U} \equiv \left(1 - w_U^N - w_U^Y\right)\theta_0 + w_U^N\left(\theta + \epsilon_N\right) + w_U^Y\left(\theta + \epsilon_S + \frac{Q}{\rho_S}\right)$$

$$\theta_I = \frac{\rho_0}{\rho_I}\theta_0 + \frac{\rho_N}{\rho_I}N + \frac{\rho_S}{\rho_I} \equiv \left(1 - w_I^N - w_I^Y\right)\theta_0 + w_I^N\left(\theta + \epsilon_N\right) + w_I^S\left(\theta + \epsilon_S\right)$$

If stocks trade at price p at date 0, informed investors will demand $x_I(p) = \rho_I\gamma^{-1}(\theta_I - rp)$ units, and behavioral investors will demand $x_U(p) = \rho_U(\theta_U - rp)$ units. The equilibrium price $p^\star$ solves $x_I + x_U + Q = 0$, or equivalently,

$$rp^\star \quad = \quad q\theta_U + (1-q)\theta_I + zQ, \tag{12}$$

where $q = \lambda\rho_U/(\lambda\rho_U + (1-\lambda)\rho_I)$ and $z = 1/\left(\lambda\rho_U + (1-\lambda)\rho_I\right)$. The approximate volatility of equilibrium returns, letting $\boldsymbol{s} = \{s_N, s_S\}$, is

$$\frac{\theta_0^2}{r^2}\mathbb{V}[R] \simeq \mathbb{V}[\theta - rp] = \mathbb{E}\left[\mathbb{V}[\theta - rp|\mathbf{s}]\right] + \mathbb{V}\left[\mathbb{E}[\theta - rp|\mathbf{s}]\right].$$

Since beliefs θ_U and θ_I are known conditional on signals $\boldsymbol{s}$, the pricing equation (12) gives $\mathbb{V}(\theta - rp|\mathbf{s}) = \mathbb{V}(\theta + zQ|\mathbf{s}) = \rho_R^{-1} + \rho_Q^{-1}z^2$. Moreover, since $\mathbb{E}\left[\theta|\boldsymbol{s}\right] = \theta_I$, we have $\mathbb{E}[\theta - rp|\mathbf{s}] = q(\theta_I - \theta_U)$. Therefore

$$\begin{aligned}\frac{\theta_0^2}{r^2}\mathbb{V}[R] &\simeq \frac{1}{\rho_I} + z^2\frac{1}{\rho_Q} + q^2\mathbb{V}\left[\theta_U - \theta_I\right] \\ &= \frac{1}{\rho_I} + z^2\frac{1}{\rho_Q} + q^2\frac{\rho_S}{\rho_I\rho_u\left(1 + \rho_S\rho_Q\right)}.\end{aligned}$$

Regarding the volatility of returns, it is possible to show that in the noiseless and noisy limits both social and news media buzz reduce volatility:

$$\lim_{\rho_Q \to 0}\frac{\partial\mathbb{V}\left[R\right]}{\partial\rho_N} < 0, \ \lim_{\rho_Q \to \infty}\frac{\partial\mathbb{V}\left[R\right]}{\partial\rho_N} < 0,$$

$$\lim_{\rho_Q \to 0}\frac{\partial\mathbb{V}\left[R\right]}{\partial\rho_S} < 0, \ \lim_{\rho_Q \to \infty}\frac{\partial\mathbb{V}\left[R\right]}{\partial\rho_S} < 0.$$

Now we consider turnover. We begin by determining the equilibrium demand by each group. Using the equilibrium condition (12), we obtain that the equilibrium demands of uninformed and informed traders, respectively, are $x_U(p^\star) = \frac{\rho_u}{\gamma}\left(q\left(\theta_U - \theta_I\right) - zQ\right)$ and $x_U(p^\star) = \frac{\rho_I}{\gamma}\left(\left(1-q\right)\left(\theta_I - \theta_U\right) - zQ\right)$. Average turnover is

$$\mathbb{E}\left[T\right]=\frac{\lambda\mathbb{E}\left[\left\Vert x_{U}\left(p^{\star}\right)\right\Vert \right]+\left(1-\lambda\right)\mathbb{E}\left[\left\Vert x_{I}\left(p^{\star}\right)\right\Vert \right]+\mathbb{E}\left[\left\Vert Q\right\Vert \right]}{2}.$$

We are not able to sign the impact of buzz on $\mathbb{E}\left[T\right]$ in general. Thus we consider its impact on uninformed and informed trades separately (the volume of noise trades $\mathbb{E}\left[\left\Vert Q\right\Vert \right]$ is independent of buzz). Note that trades are normally distributed, so that their expected absolute value is proportional to their standard deviation. In the noisy limit, it is straightforward to show that

$$\lim_{\rho_{Q}\rightarrow0}\frac{\partial st.dev.\left[\left\Vert x_{U}\left(p^{\star}\right)\right\Vert \right]}{\partial\rho_{N}}=+\infty=\lim_{\rho_{Q}\rightarrow0}\frac{\partial st.dev.\left[\left\Vert x_{I}\left(p^{\star}\right)\right\Vert \right]}{\partial\rho_{S}},$$
$$\lim_{\rho_{Q}\rightarrow0}\frac{\partial st.dev.\left[\left\Vert x_{U}\left(p^{\star}\right)\right\Vert \right]}{\partial\rho_{S}}=-\infty=\lim_{\rho_{Q}\rightarrow0}\frac{\partial st.dev.\left[\left\Vert x_{I}\left(p^{\star}\right)\right\Vert \right]}{\partial\rho_{N}}.$$

Social media buzz increases the average trade of informed investors and decreases that of uninformed traders. News media buzz has the opposite impact. The effect on uninformed traders dominates whenever their mass λ is large enough, and vice versa, which yields the result in Proposition 2.

B.2 Alternative behavioral biases

We present models in which social media is informative, but the interpretation of its content is subject to behavioral biases. Our goal is to capture some of the most common biases studied in behavioral finance. We restrict ourselves to models where "behavioral investors" have Gaussian posteriors, so their demand is linear as in the standard CARA-Gaussian model. We propose a catalog of such biases which yield the testable predictions of Proposition 3. Unless otherwise specified, we continue to use the notation of Section 3.

Investors have a common prior belief that $\theta\sim\mathcal{N}(\theta_{0},\rho_{0}^{-1})$. There are K social media signals $\boldsymbol{s}=(s_{1},...,s_{K})$, each with precision ρ_{ϵ}. We assume that news media signals, if present, are processed in line with Bayes' law by all investors, and are therefore implicit in the common prior. Rational investors form the Bayesian posterior $\theta\sim\mathcal{N}\left(\theta_{R},\rho_{R}^{-1}\right)$, where

$$\theta_{R} = \sum_{i=1}^{K}w_{i}s_{i}+\left(1-\sum_{i=1}^{K}w_{i}\right)\theta_{0}, \tag{13}$$
$$\rho_{R} = \rho_{0}+K\rho_{\epsilon}. \tag{14}$$

The Bayesian updating weights are $w_{i}=\rho_{\epsilon}/\rho_{B}$. Generically, behavioral investors will have posterior beliefs $\theta|\mathbf{s} \sim \mathcal{N}(\theta_B, \rho_R^{-1})$, where

$$\theta_B = \sum_i \hat{w}_i(s_i + \eta_i) + \left(1 - \sum_i \hat{w}_i\right)\theta_0, \tag{15}$$
$$\rho_B = \hat{\rho}_0 + \sum_i \hat{\rho}_i. \tag{16}$$

The updating rules of behavioral investors shown above can exhibit three deviations from Bayes' rule. First, the weight attributed to signals by behavioral investors ($\hat{w}_i$) may differ from the rational weights w_i. Second, the precision attributed to priors ($\hat{\rho}_0$) and each signal i ($\hat{\rho}_i$) when deriving the posterior precision can differ from the true ρ_0 and ρ_ϵ. Finally, the perception of the levels of the signals can differ from the truth by a (potentially stochastic) term η_i.

Note that Equations (5), (7) and (8) in the main text remain valid for equilibrium prices, the (approximate) volatility of returns $\mathbb{V}[R]$, and turnover $\mathbb{E}[T]$. We now characterize these quantities as a function of social media buzz K under various behavioral biases. If behavioral biases are absent ($\lambda = 0$), trading activity is zero by the "no trade theorem" (Milgrom and Stokey, 1982), and volatility is simply the posterior variance of rational traders ρ_R^{-1}, which is decreasing in buzz K.

Overconfidence

Overconfidence is commonly modeled by assuming that perceive the correct signals ($x_i = 0$) and prior variance ($\hat{\rho}_0 = \rho_0$), but believe that social media signals have precision $\hat{\rho}_i = (1 + a)\rho_\epsilon$, where $a > 0$ measures overconfidence. Thus they use the overconfident updating weights $\hat{w}_i = \hat{\rho}_i/\rho_B > w_i$.

The disagreement between behavioral and rational investors is $\theta_B - \theta_R = K(\hat{w}_i - w_i)(\theta_0 - \bar{s})$, where $\bar{s} = N^{-1}\sum_i s_i$ is the average signal. The disagreement is normally distributed with mean zero and variance $\mathbb{V}[\theta_B - \theta_R] = K^2(\hat{w}_i - w_i)^2\mathbb{V}[\bar{s}]$. The absolute value of disagreement $\|\theta_B - \theta_R\|$ has a folded normal distribution with mean $\mathbb{E}[\|\theta_B - \theta_R\|] = (2\mathbb{V}[\theta_B - \theta_R]/\pi)^{1/2}$. Substituting into (7) and (8) and differentiating yields

$$\frac{\partial \mathbb{V}[R]}{\partial K} \overset{\text{sign}}{=} a\lambda - (1 + a\lambda)^2 \frac{K\rho_\epsilon}{\rho_0} - 1$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} \overset{\text{sign}}{=} (1 - a\lambda)\frac{K\rho_\epsilon}{\rho_0} + 1.$$

This yields the predictions of Proposition 3.

Conservatism

We capture conservatism by assuming that behavioral investors correctly perceive signals ($x_i = 0$) and their precision ($\hat{\rho}_i = \rho_\epsilon$), but believe the precision of their prior to be $\hat{\rho}_0 = (1+b)\rho_0$ for $b > 0$. Now, behavioral investors use the conservative weights $\hat{w}_i = \rho_\epsilon/\rho_B < w_i$. The analysis is analogous to the case of overconfidence, and we find that buzz unambiguously decreases volatility but increases turnover:

$$\frac{\partial \mathbb{V}[R]}{\partial K} < 0,$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} > 0.$$

Rational inattention

Behavioral (rationally inattentive) traders observe social media signal s_i with cognitive noise $\eta_i \sim \mathcal{N}\left(0, \rho_{\eta,i}^{-1}\right)$, but optimally choose the $\rho_{\eta,i}$ subject to an upper bound on entropy reduction. Letting $c_i = \rho_\epsilon - \left(\rho_\epsilon^{-1} + \rho_{\eta,i}^{-1}\right)^{-1}$ be the decline in the precision of signal i due to inattention, the entropy reduction achieved by behavioral traders is determined by and increasing in the signal-to-noise ratio $\left(K\rho_\epsilon - \sum_{i=1}^K c_i\right)/\rho_0$, so the attention constraint is $\sum_i c_i \geq \underline{c}$ for an appropriate $\underline{c}$. To ensure that the attention constraint is meaningful, we assume that behavioral investors cannot infer information from prices for free.

By Theorem 1 in Peng and Xiong (2006), behavioral traders wish to maximize the posterior precision $K\rho_\epsilon - \sum_{i=1}^K c_i$, and are therefore indifferent between all choices which satisfy the binding attention constraint $\sum_i c_i = \underline{c}$. For the simplest possible exposition, we assume here that behavioral traders observe the first $k < K$ signals perfectly, but do not pay attention to the remaining $K-k$ signals ($\rho_{\eta,i} = +\infty$ for $i \leq k$, and $\rho_{\eta,i} = 0$ for $i > k$). This is exactly optimal when $\bar{c}/\rho_\epsilon$ is an integer, and a convenient approximation otherwise. Rational investors process all K signals.

Let $\bar{s}_1 = k^{-1}\sum_{i\leq k} s_i$ denote the average signal observed by behavioral investors, and $\bar{s}_2 = (K-k)^{-1}\sum_{i>k} s_i$ the average of the remaining signals. The weights placed on each signal by behavioral and rational investors are, respectively, $\hat{w}_i = \rho_\epsilon/(\rho_0 + k\rho_\epsilon)$ and $w_i = \rho_\epsilon/(\rho_0 + K\rho_\epsilon)$. Using these weights in (1) and (3) we obtain posterior means:

$$\begin{aligned}\theta_R &= w_i\left(k\bar{s}_1 + (K-k)\,\bar{s}_2\right) + (1 - Kw_i)\theta_0 \\ \theta_B &= \hat{w}_i k\bar{s}_1 + (1 - \hat{w}_i k)\theta_0.\end{aligned}$$

The disagreement $\theta_B - \theta_R$ has mean zero and variance $\mathbb{V}\left[\theta_B - \theta_R\right] = \rho_B^{-1} - \rho_R^{-1}$, implying $\mathbb{E}\left[\|\theta_B - \theta_R\|\right] = \left(2\mathbb{V}\left[\theta_B - \theta_R\right]/\pi\right)^{1/2}$. Substituting into (7) and (8) and differentiating, we find that buzz decreases the variance of returns, but increases turnover:

$$\frac{\partial \mathbb{V}[R]}{\partial K} < 0,$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} > 0.$$

Confirmation bias

Behavioral traders use the rational precisions ($\hat{\rho}_i = \rho_\epsilon$ and $\hat{\rho}_0 = \rho_0$) and rational weights $w_i = \rho_\epsilon/\rho_R$ for updating, but have an optimistic predisposition (the pessimistic case is analogous). Thus they interpret positive signals $s_i > \theta_0$ correctly, but take negative signals $s_i < \theta_0$ to be equal to their prior. The perceived signal is therefore $s_i + \eta_i$ where $\eta_i = \max\{0, \theta_0 - s_i\} \geq 0$ is the misperception due to confirmation bias. The misperception has a censored Gaussian distribution, and it is possible to show, extending the argument of Muthen (1990), that the joint moments of any two misperceptions (η_i, η_j) satisfy

$$\begin{aligned}
\mathbb{E}\left[\eta_i\right] &= \sqrt{\frac{1}{2\pi}\mathbb{V}\left[\theta_0 - s_i\right]} = \sqrt{\frac{1}{2\pi}\left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon}\right)}, \\
\mathbb{V}[\eta_i] &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon}\right)\left(\frac{1}{2} - \frac{1}{2\pi}\right), \\
Cov[\eta_i, \eta_j] &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon}\right)\left[\zeta(r)r - \frac{1}{2\pi}\left(1 - \sqrt{1 - r^2}\right)\right],
\end{aligned}$$

where $r = \rho_\epsilon/(\rho_0 + \rho_\epsilon)$ is the correlation between two signals s_i and s_j, and

$$\zeta(r) = \frac{1}{4} + \frac{1}{2\pi}\text{ArcSin}(r)$$

denotes the probability that two signals both lie below the prior. The disagreement between behavioral and rational traders is $\theta_B - \theta_R = N w_i \bar{\eta}$, where $\bar{\eta} = K^{-1}\sum_i \eta_i$ is the average misperception, and has moments

$$\begin{aligned}
\mathbb{E}\left[\theta_B - \theta_R\right] &= K w_i \mathbb{E}\left[\eta_i\right] = K w_i \sqrt{\frac{1}{2\pi}\left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon}\right)}, \\
\mathbb{V}\left[\theta_B - \theta_R\right] &= K^2 w_i^2 \left\{\frac{1}{K}\mathbb{V}\left[\eta_i\right] + \left(1 - \frac{1}{K}\right) Cov\left[\eta_i, \eta_j\right]\right\}.
\end{aligned}$$

Note further that $\mathbb{E}\left[\theta_B - \theta_R\right] = \mathbb{E}\left[\|\theta_B - \theta_R\|\right]$ since behavioral traders are weakly more optimistic than rational ones. Substituting into (7) and (8) and differentiating, we can sign the effect of buzz on volatility turnover in general, and the effect of buzz on volatility in the limiting case with a large number of signals:

$$\lim_{K\to\infty}\frac{\partial\mathbb{V}[R]}{\partial K}\overset{\text{sign}}{=}\frac{2-r}{r}\left[\zeta(r)r-\frac{1}{2\pi}\left(1-\sqrt{1-r^2}\right)\right]-\left(\frac{1}{2}-\frac{1}{2\pi}\right)-(1-r). \tag{17}$$
$$\frac{\partial\mathbb{E}[T]}{\partial K}>0.$$

To check the sign in (17), we note that the right-hand side is negative for all $r\in(0,1)$ as long as

$$1-r\geq\frac{2-r}{r}\left[\zeta(r)r-\frac{1}{2\pi}\left(1-\sqrt{1-r^2}\right)\right]-\left(\frac{1}{2}-\frac{1}{2\pi}\right)\equiv\tau(r).$$

We have $\tau(1)=0$ and $\lim_{r\to 0}\tau(r)=1/2\pi<1$, so the above holds at both boundaries of the set $(0,1)$. It is sufficient to show that $\tau'(r)\geq -1$, which rules out any crossings with $\tau(r)=1-r$ on the interior of the set. We have

$$\begin{aligned}\tau'(r)&=\frac{1-\sqrt{1-r^2}}{\pi r^2}-\frac{1}{4\pi}(\pi+2\text{ArcSin}(r))\\&\geq\frac{1}{2\pi}-\frac{1}{2}>-1,\end{aligned}$$

where the second line uses the facts that the term $\frac{1-\sqrt{1-r^2}}{\pi r^2}$ is strictly increasing in r and that $\text{ArcSin}(r)\leq\pi/2$. Thus, for large enough K, social media buzz decreases volatility and increases turnover.

B.3 Endogenous buzz

We analyze two models where buzz is endogenously determined. Unless otherwise stated, we maintain the notation of Section 3.

First, we consider the effect of disagreement on social media buzz and market activity. We assume that all traders have prior beliefs $\theta\sim\mathcal{N}\left(\theta_0,\rho_0^{-1}\right)$. News media signals, if present, are processed in line with Bayes' law by all investors, and are therefore reflected in the common prior. Then, there is a *disagreement shock* which changes the mean belief of behavioral traders to $\theta_B=\theta_0+\delta\epsilon$, where $\epsilon\sim\mathcal{N}(0,1)$ is noise (perfectly correlated across traders) and $\delta>0$ measures the extent of disagreement. Social media buzz is determined as an increasing function of disagreement and other factors u: $K=k(\delta,u)$. Social media signals are therefore uninformative and reflect only prior beliefs. Applying Equations (7) and (8) we find that return volatility and turnover between dates 1 and 2 in equilibrium satisfy

$$\begin{aligned}\frac{\theta_0^2}{r^2}\mathbb{V}[R]&\simeq\frac{1}{\rho_0}+(\lambda\delta)^2,\\\mathbb{E}[T]&=\rho_0\delta\frac{\lambda(1-\lambda)}{\gamma}\sqrt{\frac{1}{2\pi}},\end{aligned}$$

both of which are increasing in δ. An increase in δ increases both social media buzz and subsequent volatility and turnover. However, holding the measure of disagreement δ constant, volatility and turnover are independent of social media buzz.

Second, we consider the effect of prior uncertainty on the demand for news and market activity. Social and news media both respond to the demand for information, which is a function of the prior precision of beliefs ρ_0. Assume that equilibrium in the market for information is described by linear relationships determining the number of signals from news media (N) and social media (K):

$$
\begin{aligned}
N &= n_0 + n_1\rho_0 + u_N, \\
K &= k_0 + k_1\rho_0 + u_K,
\end{aligned}
$$

where u_N and u_K capture other drivers of information demand and supply (we ignore the restriction that N and K are integers for simplicity). All traders follow Bayes' law, and news and social media signals have precision ρ_N and ρ_K respectively. The volatility of returns in equilibrium satisfies

$$
\begin{aligned}
\frac{\theta_0^2}{r^2}\mathbb{V}[R] &\simeq \frac{1}{\rho_0 + N\rho_N + K\rho_K} \\
&= \frac{1}{\rho_0\left(1 + \rho_N n_1 + \rho_K k_1\right) + \rho_N\left(n_0 + u_N\right) + \rho_K\left(k_0 + u_K\right)}.
\end{aligned}
$$

Assuming that $1 + \rho_N n_1 + \rho_K k_1 > 0$, an increase in prior uncertainty (that is, a decrease in ρ_0) increases the subsequent volatility of returns. It further increases news media buzz if $n_1 > 0$ and social media buzz if $k_1 > 0$, both of which are true when the demand for information is increasing in prior uncertainty.